

Timely Throughput of Heterogeneous Wireless Networks: Fundamental Limits and Algorithms

Sina Lashgari and A. Salman Avestimehr

Abstract

The proliferation of different wireless access technologies, together with the growing number of multi-radio wireless devices suggest that the opportunistic utilization of multiple connections at the users can be an effective solution to the phenomenal growth of traffic demand in wireless networks. In this paper we consider the downlink of a wireless network with N Access Points (AP's) and M clients, where each client is connected to several out-of-band AP's, and requests delay-sensitive traffic (e.g., real-time video). We adopt the framework of Hou, Borkar, and Kumar, and study the maximum total timely throughput of the network, denoted by C_{T^3} , which is the maximum average number of packets delivered successfully before their deadline. Solving this problem is challenging since even the number of different ways of assigning packets to the AP's is N^M . We overcome the challenge by proposing a deterministic relaxation of the problem, which converts the problem to a network with deterministic delays in each link. We show that the additive gap between the capacity of the relaxed problem, denoted by C_{det} , and C_{T^3} is bounded by $2\sqrt{N(C_{\text{det}} + \frac{N}{4})}$, which is asymptotically negligible compared to C_{det} , when the network is operating at high-throughput regime. In addition, our numerical results show that the actual gap between C_{T^3} and C_{det} is in most cases much less than the worst-case gap proven analytically. Moreover, using LP rounding methods we prove that the relaxed problem can be approximated within additive gap of N . We also extend the analytical results to the case of time-varying channel states and weighted total timely throughput.

Index Terms

Wireless networks, heterogeneous networks, timely throughput, scheduling, real-time traffic, deterministic relaxation, network capacity.

S. Lashgari and A. S. Avestimehr are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY (email: sl2232@cornell.edu, avestimehr@ece.cornell.edu).

The research of A. S. Avestimehr and S. Lashgari was supported in part by Intel, Cisco, and Verizon (via the Video Aware Wireless Networks (VAWN) Program), and the U.S. Air Force award FA9550-11-1-0064.

I. INTRODUCTION

Consumer demand for data services over wireless networks has increased dramatically in recent years, fueled both by the success of online video streaming and popularity of video-friendly mobile devices like smartphones and tablets. This confluence of trends is expected to continue and lead to several fold increase in traffic over wireless networks by 2015, the majority of which is expected to be video [1]. As a result, one of the most pressing challenges in wireless networks is to find effective ways to provide high volume of top quality video traffic to smartphone users.

With the evolution of wireless networks towards heterogeneous architectures, including wireless relays and femtocells, and growing number of smart devices that can connect to several wireless technologies (e.g. 3G and WiFi), it is promising that the opportunistic utilization of heterogeneous networks (where available) can be one of the key solutions to help cope with the phenomenal growth of video demand over wireless networks. This motivates two fundamental questions: first, how much is the ultimate capacity gain from opportunistic utilization of network heterogeneity for delay-sensitive traffic? and second, what are the optimal policies that exploit network heterogeneity for delivery of delay-sensitive traffic?

In this paper, we study these questions in the downlink of a heterogeneous wireless network with N Access Points (AP's) and M clients, where each client can request data from a subset of AP's. We assume that each AP is using a distinct frequency band. Therefore, the channels from AP's to the clients are orthogonal. Moreover, we assume that all AP's are connected to each other through a Backhaul Network (see Fig. 1a), with error free links, so that we can focus on the wireless segment of the problem. We model the wireless channels from every AP to each receiver as i.i.d. packet erasure channels.

We focus on real-time video streaming applications, such as video-on-demand, video conferencing, and IPTV (TV distributed over an IP network), that require tight guarantees on the timely delivery of the packets. In particular, the packets for such applications have strict-per-packet deadline; and if a packet is not delivered successfully by its deadline, it will not be useful anymore. As a result, we focus on the notion of *timely throughput*, proposed in [2], which measures the long-term average number of “successful deliveries” (i.e., the packets that are delivered before the deadline) for each client as an analytical metric for evaluating both

throughput and QoS for delay-constrained flows.

In this framework, the time is slotted and time-slots are grouped to form intervals of length τ . For each interval every client has a packet to receive and the AP's have to decide on a scheduling policy to deliver the packets. If a packet is not delivered by the end of that interval, it gets dropped by the AP's. Timely throughput of a client is defined as the long-term average number of successful deliveries for that client. Furthermore, the total timely throughput (T^3) is the summation of timely throughputs of all the M clients in the network. Our objective is then to find the maximum achievable T^3 , which we denote by C_{T^3} , over all possible scheduling policies.

The challenge is that for each interval, even the number of different ways of assigning packets to the AP's is N^M , which grows exponentially in the number of clients (M). Therefore, we face a quite complicated combinatorial optimization problem to find C_{T^3} . To overcome this challenge, we propose a *deterministic relaxation* of the problem, which is based on converting the problem to a network with deterministic delays for each link. As we will show in Section III, the relaxed problem can be viewed as a packing problem, in which each AP turns into a bin with certain capacity and each packet turns into an object which has different sizes at different bins. The relaxed problem is then to maximize the total number of objects that can be packed in the bins, denoted by C_{det} .

Our main contribution in this paper is two-fold. First, we prove that the gap between the solutions to the original problem and its relaxed version (i.e., the gap between C_{T^3} and C_{det} is at most $2\sqrt{N(C_{\text{det}} + \frac{N}{4})}$. By noting that the number of AP's, N , is typically very small (in most cases between 2-4), the above result indicates that C_{det} is asymptotically equal to C_{T^3} as $C_{T^3} \rightarrow \infty$. Furthermore, our numerical results demonstrate that the gap between the proposed relaxed problem and the original problem is in most cases much smaller than the worst-case gap that we prove analytically. Therefore, instead of solving our main maximization problem we can solve its relaxed version, and still get a value which is very close to the optimum.

Second, we prove that the relaxed problem can be approximated in polynomial-time (with additive gap of N) using a simple LP rounding method. This approximation is appealing as N is usually limited and negligible compared to C_{det} . As a result, the solution to the relaxed problem provides a scheduling policy that provably achieves a T^3 that is within additive gap of

$N + 2\sqrt{N(C_{T^3} - \frac{3N}{4})}$ of C_{T^3} for $C_{T^3} > \frac{7N}{4}$.

We also extend our problem formulation to the case of having time-varying channels and real-time traffic, where at the beginning of each interval clients have request for variable number of packets. We show that the aforementioned results hold in these two extensions, too. Moreover, when we allow for different flows to have different weights, and for bounded ratio of the weights, denoted by ω_{max} , we show that the gap between the weighted total timely throughput, C_{w-T^3} , and the solution to the relaxed problem, C_{w-det} , is at most $2\omega_{max}\sqrt{N(C_{w-det} + \frac{N}{4})}$.

Related Work: Although there are classical results [5], [6] on scheduling clients over time-varying channels and characterizing the average delay of service, in recent years there has been increasing research on serving delay-sensitive traffic over wireless networks. This increase is due to the phenomenal increase in the volume of delay-sensitive traffic, such as video traffic. Dua et. al. [7] have considered weights and strict deadlines for packets; and if a packet is not delivered by its deadline, it causes a certain distortion equal to its weight. They have studied the problem of minimizing the total distortion and have characterized the optimal control. Agarwal and Puri [8] have considered a packet switched network where clients can get different types of service based on the amount they are willing to pay. Neely [9] has considered the problem of optimizing time averages in systems with i.i.d behavior over renewal frames. An algorithm which minimizes drift-plus-penalty ratio is developed in [9]. Shakottai and Srikant [10] have focused on minimizing the total number of expired packets, and have provided analytical results on scheduling.

However, the most related work to this paper is the work of Hou et. al in [2] in 2009, in which they have proposed a framework for jointly addressing delay, delivery ratio, and channel reliability. For a network with one AP and N clients, the timely throughput region for the set of N clients has been fully characterized in [2]; and the work has been extended to variable-bit-rate applications in [3], and time-varying channels and rate adaptation in [4]. Although in [2]- [4] they provide tractable analytical results and low-complexity scheduling policies, the analyses are done for the case of having only one AP. This paper aims to extend the results to the case of having multiple AP's, where there is an additional challenge of how to split the packets among different AP's.

The rest of the paper is organized as follows. Section II defines the notion of timely throughput, describes our network model, and explains the problem formulation. Section III describes and

formulates the relaxation of the problem, and states the main results. Section IV and Section V provide the proof of Theorem 1 and Theorem 2, respectively. Extensions to the main problem formulation are covered in Section VI. Section VII contains numerical analysis of our deterministic relaxation scheme. Finally, Section VIII concludes the paper.

II. NETWORK MODEL AND PROBLEM FORMULATION

In this section we describe our network model and precisely describe the notion of timely throughput introduced in [2]. Finally, we formulate our problem.

A. Network Model and Notion of Timely Throughput

We consider the downlink of a network with M wireless clients, denoted by Rx_1, Rx_2, \dots, Rx_M , that have packet requests, and N Access Points AP_1, AP_2, \dots, AP_N . These AP's have error-free links to the Backhaul Network (see Fig.1). In addition, we assume that time is slotted, and transmissions occur during time-slots. Furthermore, the time-slots are grouped into intervals of length τ , where the first interval contains the first τ time-slots, the second interval contains the second τ time-slots, and so on. Moreover, each AP may make one packet transmission in each time-slot.

Each AP is connected via unreliable wireless links to a subset (possibly all) of the wireless clients. These unreliable links are modeled as packet erasure channels that, for now, are assumed to be i.i.d over time, and have fixed success probabilities for each transmission. In addition, each channel is independent of other channels in the network. (In Section VI these assumptions will be relaxed to consider more general scenarios). The success probability of the channel between AP_i and Rx_j is denoted by p_{ij} , which is the probability of successful delivery of the packet of Rx_j when transmitted by AP_i during a time-slot. If there is no link between an AP and a client, we consider the success probability of the corresponding channel to be 0. Moreover, we assume that the channels do not have interference with each other.

For now we assume that at the beginning of each interval each client has request for a new packet. Right before the start of an interval, each requested packet for that interval is assigned to one of the AP's to be transmitted to its corresponding client. Furthermore, during each time-slot of an interval, each AP picks one of the packets assigned to it to transmit. At the end of that time-slot the AP will know if the packet has been successfully delivered or not. If the packet

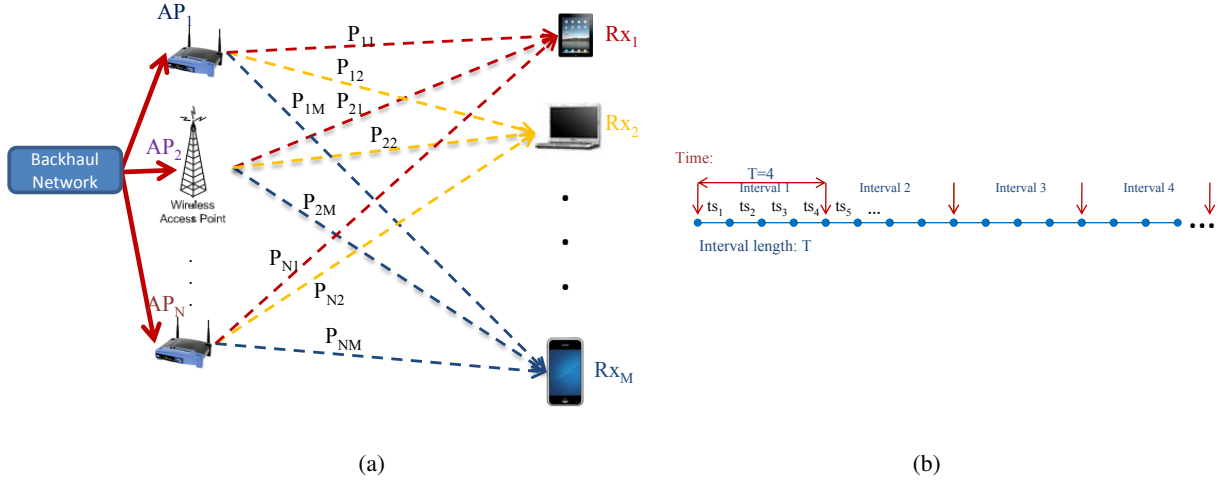


Fig. 1: Illustration of Our Network Model. Network configuration consisting of N Access points (AP's), M wireless clients, packet erasure channels from AP's to the clients, and the Backhaul network is illustrated in (a). Our model for time, in which time is slotted and time-slots are grouped to form intervals of length τ , is shown in (b). In this figure $\tau = 4$.

is successfully delivered, the AP removes that packet from its buffer and does not attempt to transmit it any more. The packets that are not delivered by the end of the interval are dropped from the AP's.

Definition 1: The decisions on how to assign the requested packets for an interval to the AP's before the start of that interval, and which packet to transmit on a time-slot by each AP are specified by a *scheduling policy*. A scheduling policy η makes the decisions causally based on the entire past history of events up to the point of decision-making. We denote the set of all possible scheduling policies by \mathcal{S} .

Definition 2: A *static scheduling policy*, denoted by η_{static} , is a scheduling policy in which each AP becomes responsible for serving packets of a fixed subset of clients for all intervals; and the packets of clients assigned to an AP are served according to a fixed order. In particular, a static scheduling policy η_{static} is fully specified by a pair $(\vec{\Pi}, \Gamma)$, in which $\vec{\Pi} = [\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_N]$, where \mathcal{I}_i 's partition the set $\{1, 2, \dots, M\}$, indicating how the packet of clients are assigned to AP's. Furthermore, Γ specifies the ordering for the packets assigned to each AP. When η_{static} is implemented, each AP is responsible for serving packet of the clients assigned to it by $\vec{\Pi}$; and

each **AP** persistently transmits a packet until it is delivered successfully, before moving on to the packet of the client with the immediate lower rank in the ordering specified by Γ .

Definition 3: A static scheduling policy is called *greedy*, and denoted by $\eta_{\text{g-static}}$, if the order of clients specified by Γ is according to the success probabilities of channels from **AP** to those clients, in decreasing order.

Assume that a particular scheduling policy η is chosen. For any interval r ($r \in \mathbb{N}$), let $\vec{N}(r, \eta) \triangleq [N_1(r, \eta), N_2(r, \eta), \dots, N_M(r, \eta)]$ denote the vector of M binary elements whose j^{th} element $N_j(r, \eta)$ is 1 if client Rx_j has successfully received a packet during the r^{th} interval, and 0 otherwise. When using scheduling policy η , the timely throughput of Rx_j , denoted by $R_j(\eta)$, is defined as

$$R_j(\eta) \triangleq \liminf_{r \rightarrow \infty} \frac{\sum_{k=1}^r N_j(k, \eta)}{r}, \quad j = 1, 2, \dots, M.^1 \quad (1)$$

In simpler words, $R_j(\eta)$ is the long-term average number of successful deliveries for the j^{th} client. Further, we denote the vector of all $R_j(\eta)$'s by $\vec{R}(\eta)$, where we have $\vec{R}(\eta) \triangleq [R_1(\eta), R_2(\eta), \dots, R_M(\eta)]$. Therefore, the capacity region for timely throughput of M clients in the network is defined as follows:

$$\mathcal{C} \triangleq \{\vec{R}(\eta) : \eta \in \mathcal{S}\}. \quad (2)$$

Finally, the total timely throughput resulting from using η , $T^3(\eta)$, is defined as

$$T^3(\eta) \triangleq \|\vec{R}(\eta)\|_1 = \sum_{j=1}^M R_j(\eta). \quad (3)$$

Therefore, $T^3(\eta)$ denotes the timely throughput of the entire network.

B. Main Problem

Our objective is to find the maximum achievable total timely throughput, denoted by C_{T^3} . More precisely, our optimization problem is

$$\text{Main Problem (MP):} \quad C_{T^3} \triangleq \sup_{\eta \in \mathcal{S}} \|\vec{R}(\eta)\|_1, \quad (4)$$

where $\|\vec{R}(\eta)\|_1 = \sum_{j=1}^M R_j(\eta)$. Later in Section VI we will consider the problem of finding the maximum weighted total timely throughput $\sum_{j=1}^M \omega_j R_j(\eta)$ and its corresponding policy η ; but for now we focus on the problem in the case that $\omega_1 = \omega_2 = \dots = \omega_M = 1$.

¹More precisely, $R_j(\eta) \triangleq \sup_{R} \liminf_{r \rightarrow \infty} \frac{\sum_{k=1}^r N_j(k, \eta)}{r} \geq R$ with probability one.

As we state later in Lemma 1 in Section IV, C_{T^3} can be achieved using a greedy static scheduling policy. Therefore, without loss of generality, the optimization in (4) can be limited to optimization over greedy static scheduling policies. So, the optimization in (4) is equivalent to finding the partition $\vec{\Pi}$ such that the corresponding $\eta_{\text{g-static}}$ maximizes $\|\vec{R}(\eta_{\text{g-static}})\|_1$. However, this is still quite challenging due to the fact that the number of possible greedy static scheduling policies to consider is N^M , which grows exponentially with M .

To overcome the aforementioned challenge, we propose a deterministic relaxation of the problem, which converts the problem to a network with deterministic delays in each link. We will show that the relaxed problem can be solved effectively, and its solution is asymptotically the same as the solution to our original problem. In the next section we will explain our deterministic relaxation approach, and then present the main result.

III. DETERMINISTIC RELAXATION AND STATEMENT OF MAIN RESULTS

In this section we first explain the intuition behind our relaxation scheme and formulate the relaxed problem. Then, we state the main results.

A. Deterministic Relaxation

Consider the system model again: we assumed channel success probability p_{ij} between AP_i and Rx_j , $i = 1, 2, \dots, N$, $j = 1, 2, \dots, M$. For now, suppose that AP_i wants to transmit a packet to client j , and that packet is the only packet assigned to AP_i for transmission. Furthermore, suppose that there is no deadline for transmitting the packet. Thus, AP_i persistently sends that packet to client j until the packet goes through. The number of time-slots expended for this packet to be delivered is a Geometric random variable G_{ij} where $\Pr(G_{ij} = k) = p_{ij}(1 - p_{ij})^{k-1}$, $k = 1, 2, \dots$. We know that $E[G_{ij}] = \frac{1}{p_{ij}}$, and without any deadline, it takes $\frac{1}{p_{ij}}$ time-slots on average for packet of Rx_j to be delivered when transmitted by AP_i .

Therefore, a memory-less erasure channel with success probability p_{ij} can be viewed as a pipe with variable delay which takes a packet from AP_i and gives it to Rx_j according to that variable delay. The probability distribution of the delay is Geometric with parameter p_{ij} .

To simplify the problem, we proposed to relax each channel into a bit pipe with deterministic delay equal to the inverse of its success probability. Therefore, for any packet of Rx_j , when assigned to AP_i for transmission, we associate a fixed size of $\frac{1}{p_{ij}}$ to that packet. This means that

each packet assigned to an AP can be viewed as an object with a size, where the size varies from one AP to another; because $\frac{1}{p_{ij}}$'s for different i 's are not necessarily the same. On the other hand, we know that each AP has τ time-slots during each interval to send the packets that are assigned to it. Therefore, we can view each AP during each interval as a bin of capacity τ . Therefore, our new problem is a packing problem; i.e., we want to see over all different assignments of objects to bins what the maximum number of objects is that we can fit in those N bins of capacity τ . We denote this maximum possible number of packed objects by C_{det} . More precisely, if we define x_{ij} as the 0 – 1 variable which equals 1 if packet of client j is assigned to AP _{i} , and 0 otherwise, then the relaxed problem can be formulated as following.

$$\begin{aligned}
 \text{Relaxed Problem (RP):} \quad C_{\text{det}} &\triangleq \max \quad \sum_{i=1}^N \sum_{j=1}^M x_{ij} \\
 \text{s.t.} \quad &\sum_{j=1}^M \frac{x_{ij}}{p_{ij}} \leq \tau \quad i = 1, 2, \dots, N \\
 &\sum_{i=1}^N x_{ij} \leq 1 \quad j = 1, 2, \dots, M \\
 &x_{ij} \in \{0, 1\}.
 \end{aligned} \tag{5}$$

B. Main Results

We now present the main results of the paper via two Theorems. Theorem 1 gives bounds on the gap between the solution to the main problem (4) and its relaxed version (5), and shows that one does not lose much if they solve the relaxed problem instead of the main problem. Furthermore, Theorem 2 proposes a performance guarantee for the polynomial-time algorithm which rounds the LP relaxation of the relaxed problem. The proofs of the two Theorems are provided in Section IV and Section V.

Theorem 1: Let C_{T^3} denote the value of the solution to our main problem in (4). Also, let C_{det} denote the value of the solution to our relaxed problem in (5). We have

$$C_{\text{det}} - 2\sqrt{N(C_{\text{det}} + \frac{N}{4})} < C_{T^3} < C_{\text{det}} + N. \tag{6}$$

Remark 1: The right part of the inequality in (6) suggests that $C_{T^3} - C_{\text{det}}$ can be no more than N . But the number of AP's N is limited and is usually around 2, 3, or 4. Therefore, as $C_{\text{det}} \rightarrow \infty$ $\frac{N}{C_{\text{det}}} \rightarrow 0$. Moreover, the left inequality in Theorem 1 suggests that $C_{\text{det}} - C_{T^3}$ can

be no more than $2\sqrt{N(C_{\text{det}} + \frac{N}{4})}$. This additive gap becomes negligible compared to C_{det} as $C_{\text{det}} \rightarrow \infty$. In addition, the inequalities in Theorem 1 imply that as $C_{T^3} \rightarrow \infty$, $C_{\text{det}} \rightarrow \infty$, too. Therefore, $\frac{C_{\text{det}}}{C_{T^3}} \rightarrow 1$, as $C_{T^3} \rightarrow \infty$. Hence, the bounds in Theorem 1 suggest the asymptotic optimality of solving C_{det} instead of C_{T^3} .

Theorem 1, basically bounds the gap between C_{T^3} and C_{det} . However, a remaining question is: if we run the system based on the greedy static scheduling policy which uses the assignment proposed by the solution to the relaxed problem, how much do we lose in terms of total timely throughput compared to C_{T^3} ? The following corollary which is proved in Appendix E addresses this question.

Corollary 1: Assume $C_{T^3} \geq \frac{7N}{4}$. Let $\vec{\Pi}_{\text{det}}$ denote the assignment of clients to AP's suggested by the solution to the relaxed problem (5), and η_{det} be the corresponding greedy static scheduling policy. Then, we have

$$C_{T^3} - N - 2\sqrt{N(C_{T^3} - \frac{3N}{4})} \leq \|\vec{R}(\eta_{\text{det}})\|_1 \leq C_{T^3}. \quad (7)$$

Remark 2: As we prove in Appendix A the upper bound given in the right inequality in Theorem 1 is tight. Furthermore, the lower bound given in the left inequality of Theorem 1 is tight in terms of order, i.e., there exists a network configuration and a positive constant k for which $C_{\text{det}} - C_{T^3} > k\sqrt{NC_{\text{det}}}$.

The proof of tightness of the bounds given by Theorem 1 can be found in Appendix A.

Remark 3: The bounds in Theorem 1 are worst-case bounds, and via numerical analysis we observe that the gap between the original problem and its relaxed version is in most cases much smaller. Therefore, the solution to the relaxed problem tracks the solution to the main problem very well, even for a limited number of clients. To illustrate this, consider the network configuration in Figure 2a. In this network there are two AP's with the same coverage radius $\frac{1}{3}$, and 10 clients which are uniformly and randomly located in the coverage area of the two AP's. The erasure probability of the channel between a client and an AP is proportional to the distance. Furthermore, the packets' deadline, τ , is assumed to be 15. For 30 different realizations of this network, C_{T^3} and C_{det} have been calculated, and plotted in Figure 2b (more detailed numerical results are provided in Section VII).

The numerical results suggest that even for small-scale networks C_{det} is usually very close to C_{T^3} . Therefore, rather than solving the main problem, we can solve the relaxed problem and

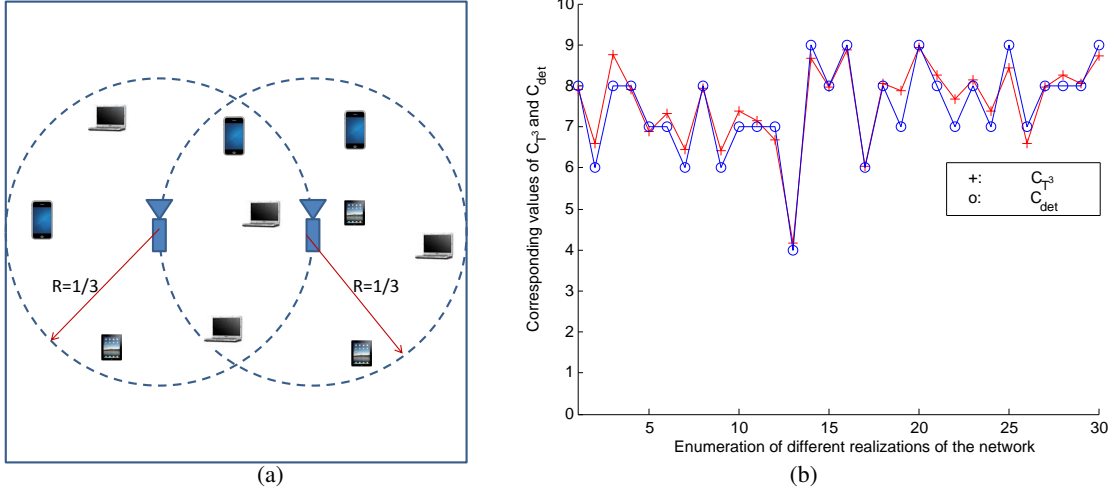


Fig. 2: Numerical analysis for the gap between C_{T^3} and C_{det} for the case of two AP's with coverage radius $\frac{1}{3}$, 10 wireless clients, and intervals of length $\tau = 15$. (a) illustrates the network configuration, where erasure probability of a channel is proportional to the distance between the AP and the corresponding receiver. (b) demonstrates the numerical results for the gap for 30 different realizations of the network, where each realization is constructed from a random and uniform location of clients in the network. Each '+' indicates the value of C_{T^3} for each realization, while 'o' indicates the value of C_{det} for the same realization.

still get a value as the solution which is very close to the solution for the main problem.

So far, we have shown by Theorem 1 that by considering the relaxed problem (RP) we do not lose much in terms of total timely throughput capacity. Nevertheless, in order for the relaxation to be useful there should be a way to solve the relaxed problem efficiently. The next Theorem, which is proved in Section V, demonstrates that the relaxed problem can be approximated efficiently.

Theorem 2: Suppose that \mathbf{x}^{*tr} is a basic optimal solution to the LP relaxation of (5). We have

$$C_{det} - \sum_{i=1}^N \sum_{j=1}^M \lfloor \mathbf{x}_{ij}^{*tr} \rfloor \leq N. \quad (8)$$

Remark 4: According to Theorem 2 if we find a basic optimal solution to LP relaxation of (5), and then round down the solution to get integral values, the result will deviate from the

optimal solution by at most N . Note that N is typically very small (in most cases between 2-4); and therefore, this algorithm performs well in approximating the optimal solution of the relaxed problem (5).

Remark 5: The relaxed problem in (5) is a special case of the well-known Maximum Generalized Assignment Problem (GAP). There is a large body of literature on GAP; and its special cases capture many combinatorial optimization problems, and have several applications in computer science and operations research. GAP is APX-hard [14], meaning that there is no polynomial-time algorithm with constant-gap performance guarantee for it. However, there are several approximation algorithms for GAP, including [14], [15]. In particular, Chekuri and Khanna [14], based on a modification of the work by Shmoys and Tardos [13], have come up with a 2-approximation algorithm for GAP; and Fleischer et. al. [15] have proposed an LP-based $\frac{e}{e-1}$ -approximation algorithm. The performance guarantees in the literature are concerned with multiplicative gap. However, our result in Theorem 2 suggests an additive gap performance guarantee of N for the special case of GAP presented in (5). Since N (the number of access points) is typically very small, this provides a tighter approximation guarantee for our problem of interest.

IV. ANALYSIS OF APPROXIMATION GAP (PROOF OF THEOREM 1)

In this section we prove Theorem 1. To this aim, we first state the following lemma which is proved in Appendix B.

Lemma 1: C_{T^3} can be achieved using a greedy static scheduling policy.

We now use Lemma 1 in order to prove the right part of the inequality in Theorem 1.

A. Proof of $C_{T^3} < C_{det} + N$:

By Lemma 1 it is sufficient to prove that for any greedy static scheduling policy $\eta_{g-static}$ we have

$$T^3(\eta_{g-static}) < C_{det} + N. \quad (9)$$

Suppose an arbitrary greedy static scheduling policy $\eta_{g-static}$ with the corresponding partition $\vec{\Pi}_{g-static} = [\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_N]$ and ordering $\Gamma_{g-static}$ is implemented. By (3) we know that

$$T^3(\eta_{g-static}) = \sum_{j=1}^M R_j(\eta_{g-static}). \quad (10)$$

On the other hand, by (1) we know that

$$R_j(\eta_{\text{g-static}}) = \liminf_{r \rightarrow \infty} \frac{\sum_{k=1}^r N_j(k, \eta_{\text{g-static}})}{r}, \quad j \in \{1, 2, \dots, M\}. \quad (11)$$

Let Y_i denote the random variable indicating the number of successful deliveries by \mathbf{AP}_i during one interval, when $\eta_{\text{g-static}}$ is implemented, $i = 1, 2, \dots, N$. More precisely,

$$Y_i \triangleq \sum_{j \in \mathcal{I}_i} N_j(1, \eta_{\text{g-static}}), \quad i = 1, 2, \dots, N. \quad (12)$$

Since a greedy static scheduling policy is implemented and channels are i.i.d over time, by LLN we have

$$E[Y_i] = \sum_{j \in \mathcal{I}_i} \liminf_{r \rightarrow \infty} \frac{\sum_{k=1}^r N_j(k, \eta_{\text{g-static}})}{r} = \sum_{j \in \mathcal{I}_i} R_j(\eta_{\text{g-static}}), \quad i = 1, 2, \dots, N. \quad (13)$$

Therefore,

$$\sum_{i=1}^N E[Y_i] = \sum_{i=1}^N \sum_{j \in \mathcal{I}_i} R_j(\eta_{\text{g-static}}) = \sum_{j=1}^M R_j(\eta_{\text{g-static}}) = \mathsf{T}^3(\eta_{\text{g-static}}). \quad (14)$$

Define $q_i \triangleq |\mathcal{I}_i|$, and denote the enumeration of clients assigned to \mathbf{AP}_i by $\{\mathcal{I}_i(1), \mathcal{I}_i(2), \dots, \mathcal{I}_i(q_i)\}$, where the enumeration is according to the channel success probabilities of different clients in \mathcal{I}_i . Let G_{ij} be a geometric random variable with parameter p_{ij} , $i = 1, 2, \dots, N$, $j = 1, 2, \dots, M$.

Then, it is easy to see that

$$Y_i = \max_k \quad k \quad \text{s.t.} \quad \sum_{j=1}^k G_{i\mathcal{I}_i(j)} \leq \tau, \quad i \in \{1, 2, \dots, N\}, k \leq q_i, \quad (15)$$

since $\eta_{\text{g-static}}$ persistently sends a packet until it is delivered, or the interval is over. Let us also define

$$l_i \triangleq \max_{\hat{l}} \quad \hat{l} \quad \text{s.t.} \quad \sum_{j=1}^{\hat{l}} 1/p_{i\mathcal{I}_i(j)} \leq \tau, \quad \hat{l} \leq q_i. \quad (16)$$

Therefore, l_i is the maximum number of objects that fit into bin of capacity τ when the channels are relaxed and clients in \mathcal{I}_i are assigned to \mathbf{AP}_i . The following lemma (for which the proof is provided in Appendix C) relates l_i to Y_i .

Lemma 2: Let $\tau \in \mathbb{N}$ and G_1, G_2, \dots, G_q be independent geometric random variables with parameters p_1, p_2, \dots, p_q respectively, such that $1 \geq p_1 \geq p_2 \geq \dots \geq p_q \geq 0$. Also define

$$l \triangleq \max_{\hat{l}} \quad \hat{l} \quad \text{s.t.} \quad \sum_{i=1}^{\hat{l}} 1/p_i \leq \tau \quad (17)$$

and

$$Y \triangleq \max i \quad s.t. \quad \sum_{j=1}^i G_j \leq \tau, \quad i \in \{1, 2, \dots, q\}. \quad (18)$$

Then, we have

$$E[Y] < l + 1. \quad (19)$$

Hence, by Lemma 2 we have

$$\sum_{i=1}^N E[Y_i] < \sum_{i=1}^N (l_i + 1). \quad (20)$$

Therefore, by (14) and (90) we have

$$\mathsf{T}^3(\eta_{\text{g-static}}) < \sum_{i=1}^N l_i + N. \quad (21)$$

But, $\sum_{i=1}^N l_i$ is the value of the objective function in (5) for a feasible solution. Therefore,

$$\sum_{i=1}^N l_i \leq C_{\text{det}}. \quad (22)$$

Putting (91) and (92) together we get

$$\mathsf{T}^3(\eta_{\text{g-static}}) < C_{\text{det}} + N. \quad (23)$$

Hence the proof of the right inequality in Theorem 1 is complete.

B. Proof of $C_{\text{det}} - 2\sqrt{N(C_{\text{det}} + \frac{N}{4})} < C_{\mathsf{T}^3}$

Consider the assignment proposed by the solution to the relaxed problem in (5), where the clients that are not assigned to any AP for transmission are assigned to AP's arbitrarily. Let $\vec{\Pi}_{\text{g-static}}^{\text{det}} = [\mathcal{I}_1^{\text{det}}, \mathcal{I}_2^{\text{det}}, \dots, \mathcal{I}_N^{\text{det}}]$ denote the resulting partition, and also let $\eta_{\text{g-static}}^{\text{det}}$ denote the corresponding greedy static scheduling policy. Therefore, we have

$$\mathsf{T}^3(\eta_{\text{g-static}}^{\text{det}}) \leq C_{\mathsf{T}^3}.$$

So, it is sufficient to prove that $C_{\text{det}} - 2\sqrt{N(C_{\text{det}} + \frac{N}{4})} < \mathsf{T}^3(\eta_{\text{g-static}}^{\text{det}})$. Let Y_i^{det} denote the random variable indicating the number of successful deliveries by AP_{*i*} during one interval, when $\eta_{\text{g-static}}^{\text{det}}$ is implemented, $i = 1, 2, \dots, N$. With the same argument as in part A we have

$$\mathsf{T}^3(\eta_{\text{g-static}}^{\text{det}}) = \sum_{i=1}^N E[Y_i^{\text{det}}].$$

Therefore, it is sufficient to prove that

$$C_{\text{det}} - 2\sqrt{N(C_{\text{det}} + \frac{N}{4})} < \sum_{i=1}^N E[Y_i^{\text{det}}].$$

Define $q_i = |\mathcal{I}_i^{\text{det}}|$; and denote the enumeration of clients assigned to AP_i by $\{\mathcal{I}_i(1), \mathcal{I}_i(2), \dots, \mathcal{I}_i(q_i)\}$, where the enumeration is according to the channel success probabilities of different clients in \mathcal{I}_i . Further, let G_{ij} be a geometric random variable with parameter p_{ij} , $i = 1, 2, \dots, N$, $j = 1, 2, \dots, M$. Then, it is easy to see that

$$Y_i^{\text{det}} = \max_k \quad k \quad \text{s.t.} \quad \sum_{j=1}^k G_{i\mathcal{I}_i(j)^{\text{det}}} \leq \tau, \quad i \in \{1, 2, \dots, N\}, k \leq q_i, \quad (24)$$

since $\eta_{\text{wg-static}}^{\text{det}}$ persistently sends a packet until it is delivered, or the interval is over. Also define

$$l_i^{\text{det}} \triangleq \max_{\hat{l}} \quad \hat{l} \quad \text{s.t.} \quad \sum_{j=1}^{\hat{l}} 1/p_{i\mathcal{I}_i^{\text{det}}(j)} \leq \tau, \quad \hat{l} \leq q_i.$$

Therefore, l_i^{det} is the maximum number of objects that fit into a bin of capacity τ when the channels are relaxed and clients in $\mathcal{I}_i^{\text{det}}$ are assigned to AP_i . The following lemma (which is proved in Appendix D) relates l_i^{det} to Y_i^{det} .

Lemma 3: Let $\tau \in \mathbb{N}$ and G_1, G_2, \dots, G_q be independent geometric random variables with parameters p_1, p_2, \dots, p_q respectively, such that $1 \geq p_1 \geq p_2 \geq \dots \geq p_q \geq 0$. Also define

$$l \triangleq \max_{\hat{l}} \quad \hat{l} \quad \text{s.t.} \quad \sum_{i=1}^{\hat{l}} 1/p_i \leq \tau$$

and

$$Y \triangleq \max_i \quad i \quad \text{s.t.} \quad \sum_{j=1}^i G_j \leq \tau, \quad i \in \{1, 2, \dots, q\}.$$

Then, we have

$$l - 2\sqrt{l + \frac{1}{4}} < E[Y]. \quad (25)$$

By Lemma 3 we have

$$\sum_{i=1}^N l_i^{\text{det}} - 2 \sum_{i=1}^N \sqrt{l_i^{\text{det}} + \frac{1}{4}} < \sum_{i=1}^N E[Y_i^{\text{det}}]. \quad (26)$$

According to Cauchy-Schwarz inequality we have

$$\sum_{i=1}^N \sqrt{l_i^{\text{det}} + \frac{1}{4}} \leq \sqrt{N \left(\sum_{i=1}^N l_i^{\text{det}} + \frac{N}{4} \right)}. \quad (27)$$

On the other hand, note that

$$C_{\det} = \sum_{i=1}^N l_i^{\det}. \quad (28)$$

Putting (26), (27), and (28) together we get

$$C_{\det} - 2\sqrt{N(C_{\det} + \frac{N}{4})} < C_T^3. \quad (29)$$

Hence, the left inequality of Theorem 1 is proved and the proof of Theorem 1 is complete.

V. EFFICIENT SOLUTION TO THE RELAXED PROBLEM (PROOF OF THEOREM 2)

In this section we show that the approximation proposed by rounding down a basic optimal solution to LP-relaxation of RP in (5) is at most N away from C_{\det} . Note that RP is a mixed integer linear program. Linear relaxation of such a problem replaces the integrality constraints with non-negativity constraints; i.e., it replaces the constraint $x_{ij} \in \{0, 1\}$ with $x_{ij} \geq 0$ for $i = 1, 2, \dots, N$, $j = 1, 2, \dots, M$. Let LR-RP denote the linear relaxation of RP. Any feasible solution to LR-RP can be denoted by an N -by- M matrix $\mathbf{x} = [x_{ij}]_{N \times M}$ such that each element of \mathbf{x} , x_{ij} , is the fraction of the j^{th} object which is assigned to the i^{th} bin; if $x_{ij} = 1$, then the j^{th} object is fully fit into the i^{th} bin; but if x_{ij} is a fractional value, then the j^{th} object is partially assigned to the i^{th} bin.

Note that finding a basic optimal solution to a linear program is straightforward, and is discussed in [18]. So, let $\mathbf{x}^{*tr} = [x_{ij}^{*tr}]_{N \times M}$ denote a basic optimal solution to LR-RP with objective value V^* ; i.e., $V^* = \sum_{i=1}^N \sum_{j=1}^M x_{ij}^{*tr}$. Define K as the number of objects that are not assigned to any of the bins (not even partially), by the solution proposed by \mathbf{x}^{*tr} . Therefore, $V^* \leq M - K$; and by noting that $C_{\det} \leq V^*$ (due to optimality of \mathbf{x}^{*tr}), we have

$$C_{\det} \leq M - K. \quad (30)$$

Moreover, suppose that there are Z objects in the solution proposed by \mathbf{x}^{*tr} each of which is associated with at least one fractional value x_{ij}^{*tr} . Therefore,

$$\sum_{i=1}^N \sum_{j=1}^M \lfloor x_{ij}^{*tr} \rfloor = M - K - Z. \quad (31)$$

By putting (30) and (31) together we get

$$C_{\det} - \sum_{i=1}^N \sum_{j=1}^M \lfloor x_{ij}^{*tr} \rfloor \leq Z. \quad (32)$$

Now, it is sufficient to show that $Z \leq N$ in order to complete the proof of Theorem 2. To do so, we convert RP to an instance of Min-GAP, which is formulated as following, and then use the result in [16], [17].

$$\begin{aligned} \text{Min-GAP:} \quad & \min \quad \sum_{i=1}^N \sum_{j=1}^M c_{ij} x_{ij} \\ & s.t. \quad \sum_{j=1}^M s_{ij} x_{ij} \leq b \quad i = 1, 2, \dots, N \\ & \quad \sum_{i=1}^N x_{ij} = 1 \quad j = 1, 2, \dots, M \\ & \quad x_{ij} \in \{0, 1\} \quad i = 1, 2, \dots, N \text{ and } j = 1, 2, \dots, M. \end{aligned} \quad (33)$$

The conversion is done as following; and it is similar to that of Section 3.2 in [14]:

- Set $s_{ij} = \frac{1}{p_{ij}}$ if $p_{ij} > 0$, and $s_{ij} = T + 1$ if $p_{ij} = 0$.
- Set $b = \tau$.
- Set $c_{ij} = 1$ for $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, M$.
- Add an additional bin of capacity τ to ensure the existence of a feasible solution. For this bin, the $(N + 1)^{st}$ bin, set $s_{N+1j} = 0$ and $c_{N+1j} = M$ for $j = 1, 2, \dots, M$.

Therefore, the converted problem is as follows.

$$\begin{aligned} \text{Converted Problem (CP):} \quad & \min \quad \sum_{i=1}^N \sum_{j=1}^M x_{ij} + \sum_{j=1}^M M x_{(N+1)j} \\ & s.t. \quad \sum_{j=1}^M s_{ij} x_{ij} \leq \tau \quad i = 1, 2, \dots, N + 1 \\ & \quad \sum_{i=1}^{N+1} x_{ij} = 1 \quad j = 1, 2, \dots, M \\ & \quad x_{ij} \in \{0, 1\} \quad i = 1, 2, \dots, N + 1 \text{ and } j = 1, 2, \dots, M. \end{aligned} \quad (34)$$

Let LR-CP denote the linear relaxation of the converted problem in (34). The following lemma, which is proved later in this section, connects basic optimal solutions of LR-RP and LR-CP.

Lemma 4: Suppose that $\mathbf{x}^{*tr} = [x_{ij}^{*tr}]_{N \times M}$ is a basic solution to LR-RP. Further, consider $\mathbf{x}^* = [x_{ij}^*]_{(N+1) \times M}$, where

- 1) $\forall i \in \{1, 2, \dots, N\}$ and $\forall j \in \{1, 2, \dots, M\}$ $x_{ij}^* = x_{ij}^{*tr}$.
- 2) $\forall j \in \{1, 2, \dots, M\}$ $x_{N+1j}^* = 1 - \sum_{i=1}^N x_{ij}^{*tr}$.

Then, \mathbf{x}^* is a basic solution to LR-CP.

Consider the solution \mathbf{x}^* to LR-CP with the same construction as described in Lemma 4. By Lemma 4, \mathbf{x}^* is a basic solution to LR-CP. Therefore, by considering the fact that there are also exactly Z partially assigned objects in the assignment proposed by \mathbf{x}^* , and using Theorem 1 of [16], [17], we get

$$Z \leq \text{Number of bins that are full.} \quad (35)$$

Since the $(N+1)^{st}$ bin is never used up to its capacity, by (35) we have

$$Z \leq N. \quad (36)$$

Hence, by (32) and (36) we get

$$C_{\det} - \sum_{i=1}^N \sum_{j=1}^M \lfloor x_{ij}^{*tr} \rfloor \leq N. \quad (37)$$

Therefore, the proof of Theorem 2 is complete. We now prove Lemma 4.

Proof: It is easy to check that \mathbf{x}^* is a feasible solution to LR-CP. Consider the feasible region for x_{ij} 's, $1 \leq i \leq N, 1 \leq j \leq M$, in LR-RP and LR-CP; and denote the corresponding regions by \mathcal{R}_1 and \mathcal{R}_2 , respectively. Note that each point inside \mathcal{R}_1 corresponds to a feasible solution for LR-RP; and each point inside \mathcal{R}_2 corresponds to a feasible solution for the first N bins of LR-CP. It is then straightforward to check that $\mathcal{R}_1 = \mathcal{R}_2$. In addition, any basic solution to LR-RP is an extreme point of the polytope defined by \mathcal{R}_1 . Now suppose that \mathbf{x}^* is *not* a basic solution to LR-CP. Then, there exists a constant $0 < \alpha < 1$, and two non-identical matrices $\mathbf{w} = [w_{ij}]_{(N+1) \times M}, \mathbf{y} = [y_{ij}]_{(N+1) \times M}$ such that

$$x_{ij}^* = \alpha w_{ij} + (1 - \alpha) y_{ij}, \quad i = 1, 2, \dots, N+1 \quad j = 1, 2, \dots, M. \quad (38)$$

Therefore, since $x_{ij}^{*tr} = x_{ij}^*$ for $1 \leq i \leq N, 1 \leq j \leq M$, \mathbf{x}^{*tr} can be written as a convex combination of two distinct points inside \mathcal{R}_1 , which is contrary to the fact that \mathbf{x}^{*tr} is the extreme point of the polytope \mathcal{R}_1 . Hence, the assumption that \mathbf{x}^* is not a basic solution to LR-CP is incorrect, and the proof is complete. ■

Corollary 2: Suppose we choose a basic optimal solution to the LP relaxation of (5), denoted by \mathbf{x}^{*tr} , and round down the solution to get integral values. Let $\vec{\Pi}_{det}^{apx}$ denote the assignment suggested by the resulting integral values; and let η_{det}^{apx} denote the corresponding greedy static scheduling policy.. For $C_{T^3} > \frac{11N}{4}$ we have

$$C_{T^3} - 2N - 2\sqrt{N(C_{T^3} - \frac{7N}{4})} \leq \|\vec{R}(\eta_{det}^{apx})\|_1 \leq C_{T^3}. \quad (39)$$

Proof: Let $C_{det}^{apx} \triangleq \sum_{i=1}^N \sum_{j=1}^M \lfloor x_{ij}^{*tr} \rfloor$ denote the resulting objective value of the rounded down basic optimal solution to LR-RP. According to Theorem 1 and Theorem 2, $C_{det}^{apx} \geq C_{det} - N \geq C_{T^3} - 2N$. Therefore, by using the similar argument as the one in Corollary 1 the proof will be complete. ■

VI. EXTENSIONS

In this section we investigate two important extensions to our main problem formulation. First, we extend our model to allow for time-varying channels, where channels' statistics change over time. In addition, we consider real-time traffic pattern in which for each interval a client has request for a variable number of packets (possibly zero). Second, we investigate the problem of maximizing weighted total timely throughput. We present similar results as the main result in Theorem 1 for both extensions.

A. Time-Varying Channels and Real-Time Traffic

So far, we have considered a fixed packet generation for each interval, meaning that we have assumed that at the beginning of each interval each client has request for exactly one packet. This assumption can be modified by considering a time-varying packet generation pattern, in which for every interval, each client might have request for no packets, or it might have request for multiple packets. In addition, the number of packets requested by clients for one interval might depend on the number of packets requested for other intervals.

Furthermore, so far we have assumed that channel reliabilities are static, and channel success probabilities do not change from time to time. But, this model can be generalized to include time-varying channels whose statistical behaviors are not necessarily independent of one another.

We capture the above two generalizations by considering an irreducible Finite-State Markov Chain (FSMC), in which each state jointly specifies the subset of clients that have packet request,

and the number of their requested packets, as well as the channel states for different channels during an interval. When a new interval begins, the Markov Chain might change its state, and in this case, packets for a new subset of clients are requested, and the channel reliabilities change. By considering this FSMC we allow for variable-bit-rate traffic, as well as variations in channel reliabilities over time.

Let us denote the set of all possible states of the FSMC by \mathcal{C} . Each state $\lambda \in \mathcal{C}$ specifies a pair $(\vec{B}(\lambda), \mathbf{P}(\lambda))$, where $\vec{B}(\lambda) \triangleq [B_1(\lambda), B_2(\lambda), \dots, B_M(\lambda)]$, such that $B_j(\lambda)$ is the number of the packets requested by client j , and $\mathbf{P}(\lambda)$ is an $N * M$ matrix that contains channel success probabilities. It is assumed that channel success probabilities remain the same during each interval, and are known to the AP's at the beginning of each interval.

Our objective is again to find C_{T^3} . We use a similar argument as the one in [4] for extensions to time-varying channels and variable-bit-rate traffic. In particular, we decompose the set of intervals into different subsets, where each subset contains the intervals that are in the same state of the FSMC.

Consider all the intervals for which the system is at state λ . For those intervals we convert our problem to an instance of the problem described in Section II. More particularly, for the system described by state λ , we ignore all the clients that do not have packet request. Furthermore, for any $j \in \{1, 2, \dots, M\}$ where $B_j(\lambda) > 1$ we consider $B_j(\lambda) - 1$ virtual clients, such that the channel between AP_i and each of those virtual clients would have success probability $P_{ij}(\lambda)$. This means that these virtual clients are exact copies of Rx_j . Consequently, for the intervals for which the system is at state λ the problem becomes the same as the one described in Section II. With the same argument as in proof of Theorem 1, there exists a fixed assignment $\vec{\Pi}(\lambda)$, which if used together with its corresponding optimal ordering for such intervals, achieves the optimal T^3 for those intervals. We denote this optimal T^3 by $C_{T^3}(\lambda)$. In addition, let $C_{\text{det}}(\lambda)$ denote the solution to the relaxed problem when the system is at state λ . For any state $\lambda \in \mathcal{C}$, with the same argument as in the proof of Theorem 1, we have

$$C_{\text{det}}(\lambda) - 2\sqrt{N(C_{\text{det}}(\lambda) + \frac{N}{4})} < C_{T^3}(\lambda) < C_{\text{det}}(\lambda) + N. \quad (40)$$

Now, let π_λ denote the steady state probability of state λ . Therefore,

$$C_{T^3} = \sum_{\lambda \in \mathcal{C}} \pi_\lambda C_{T^3}(\lambda)$$

$$C_{\text{det}} = \sum_{\lambda \in \mathcal{C}} \pi_\lambda C_{\text{det}}(\lambda).$$

Hence,

$$C_{\text{det}} - 2 \sum_{\lambda \in \mathcal{C}} \pi_\lambda \sqrt{N(C_{\text{det}}(\lambda) + \frac{N}{4})} < C_{T^3} < C_{\text{det}} + N. \quad (41)$$

On the other hand, by using Cauchy-Schwarz inequality we have

$$\begin{aligned} & \sum_{\lambda \in \mathcal{C}} \pi_\lambda \sqrt{N(C_{\text{det}}(\lambda) + \frac{N}{4})} \\ & \leq \sqrt{\sum_{\lambda \in \mathcal{C}} \pi_\lambda} \sqrt{\sum_{\lambda \in \mathcal{C}} N \pi_\lambda (C_{\text{det}}(\lambda) + \frac{N}{4})} \\ & = \sqrt{N(C_{\text{det}} + \frac{N}{4})}. \end{aligned} \quad (42)$$

Putting (41) and (42) together we get

$$C_{\text{det}} - 2\sqrt{N(C_{\text{det}} + \frac{N}{4})} < C_{T^3} < C_{\text{det}} + N, \quad (43)$$

which is the same as the result in Theorem 1.

Theorem 3: For the network model described in Section II consider the extension to time-varying channels and real-time traffic, modeled by the FSMC described in Section VI, where each state of FSMC captures both the success probability of channels and the number of packets for each client during an interval. We have

$$C_{\text{det}} - 2\sqrt{N(C_{\text{det}} + \frac{N}{4})} < C_{T^3} < C_{\text{det}} + N. \quad (44)$$

B. Weighted Total Timely Throughput

In Section II we considered the same importance for all the flows in the network; and our objective was to maximize T^3 . However, it might be the case that in a network some of the flows are more important than the others. Therefore, those flows should somehow be prioritized accordingly.

In this section the formulation remains the same as the one described in Section II, except the objective function, which rather than maximizing T^3 , maximizes a weighted average of timely

throughputs. In particular, weighted total timely throughput of the scheduling policy η , $w\text{-}\mathsf{T}^3(\eta)$, is defined as

$$w\text{-}\mathsf{T}^3(\eta) \triangleq \sum_{j=1}^M \omega_j R_j(\eta), \quad (45)$$

where ω_j 's are arbitrary weights greater than 1 ($j = 1, 2, \dots, M$).

Our objective is to find

$$C_{w\text{-}\mathsf{T}^3} \triangleq \sup_{\eta \in \mathcal{S}} w\text{-}\mathsf{T}^3(\eta). \quad (46)$$

For this extension of the problem we again propose the channel relaxation which results in a new integer program. This integer program is again a GAP. The formulation of the relaxed problem is as follows:

$$\begin{aligned} C_{w\text{-det}} \triangleq \quad & \max \sum_{i=1}^N \sum_{j=1}^M x_{ij} \omega_j \\ \text{s.t.} \quad & \sum_{j=1}^M \frac{x_{ij}}{p_{ij}} \leq \tau \quad i = 1, 2, \dots, N \\ & \sum_{i=1}^N x_{ij} \leq 1 \quad j = 1, 2, \dots, M \\ & x_{ij} \in \{0, 1\} \quad i = 1, 2, \dots, N \quad j = 1, 2, \dots, M. \end{aligned} \quad (47)$$

The following theorem, which is proved in Appendix F, states that the value of the solution to (46) is asymptotically the same as the value of the solution to (47) for $C_{w\text{-}\mathsf{T}^3} \rightarrow \infty$ (or equivalently $C_{w\text{-det}} \rightarrow \infty$).

Theorem 4: Let $C_{w\text{-}\mathsf{T}^3}$ denote the value of the solution to (46). Further, let $C_{w\text{-det}}$ denote the value of the solution to (47). Then,

$$C_{w\text{-det}} - 2\omega_{max} \sqrt{N(C_{w\text{-det}} + \frac{N}{4})} < C_{w\text{-}\mathsf{T}^3} < C_{w\text{-det}} + N\omega_{max}, \quad (48)$$

where $\omega_{max} = \max\{\omega_1, \omega_2, \dots, \omega_M\}$.

VII. NUMERICAL ANALYSIS

In this section we provide numerical analyses for our deterministic relaxation scheme. So, we consider a wireless network with two Access Points, and several wireless clients that are

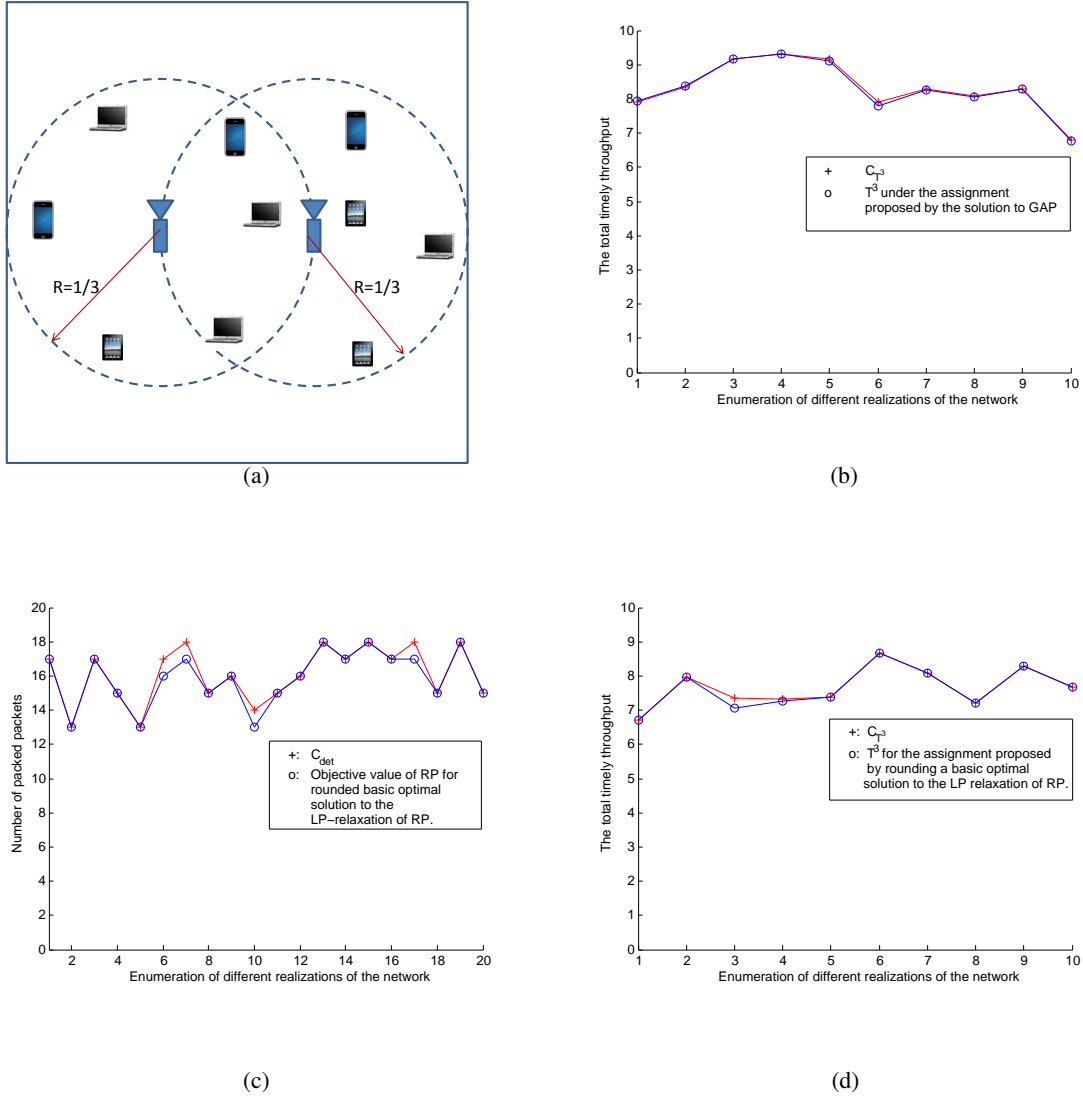


Fig. 3: Numerical Results. (a) illustrates the network configuration in which there are two AP's with coverage radius $\frac{1}{3}$ each, M randomly and uniformly located clients in the coverage area of the AP's, and channel erasure probabilities proportional to the distances. (b) compares C_{T^3} with the T^3 which is resulted from the assignment proposed by (5) for the case of $M = 10, \tau = 15$ and 10 different realizations of the network. '+' in (b) indicates the value of C_{T^3} for each realization, while 'o' indicates the value of $T^3(\eta_{det})$. (c) compares C_{det} (denoted by '+') with the objective value of the rounded basic optimal solution (denoted by 'o') for the case of $M = 20, \tau = 30$ and 20 different realizations of the network. Finally, (d) compares C_{T^3} (denoted by '+') with the T^3 resulted from the assignment proposed by rounding the basic optimal solution (denoted by 'o') for the case of $M = 10, \tau = 15$ and 10 different realizations of the network.

uniformly and randomly located in the network (see Figure 2a). The channel from every AP to every wireless client is an erasure channel with erasure probability which is proportional to the distance between the AP and the client. The distances in the network are normalized, and we assume that the AP's have the same coverage radius $R = \frac{1}{3}$. Therefore, the channel erasure probability is 1 for the channel between an AP and a client which is located at the distance $R = \frac{1}{3}$ from it. Furthermore, the distance between the two AP's is $\frac{1}{3}$.

Figure 3b corresponds to the case where $M = 10$ and $\tau = 15$. In each realization 10 clients are randomly located in the network. For each realization C_{T^3} is calculated. Then, the corresponding relaxed problem is solved, and the network is run for 10000 intervals under the assignments proposed by its deterministic relaxation solution. Fig 3b shows the comparison between the two for 30 different realizations of the network.

Figure 3c demonstrates how our proposed LP-rounding algorithm performs compared to C_{det} . We consider $M = 20$ and $\tau = 30$. 30 different realizations of network are considered. For each realization C_{det} is found. Then, we find the number of packets that are completely assigned to a bin in a basic optimal solution to the LP relaxation of (5). The result confirms the fact that our proposed algorithm performs well. The performance improves as we have larger number of clients in the network.

Figure 3d shows how far our T^3 will be if we use our rounding approximation as the assignment strategy for the packets, and run the network for 10000 intervals according to that assignment. In this case we have considered 10 clients and have looked at 10 instances of the network. τ is assumed to be 15 in this case.

VIII. CONCLUSION

In this work we investigated the improvement by utilizing network heterogeneity in order to enhance the timely throughput of a wireless network. In particular, we studied the problem of maximizing total timely throughput of the downlink of a wireless network with N Access points and M clients, where each client might have access to several Access points. This problem is challenging to attack directly. However, we proposed a deterministic relaxation of the problem which is based on converting the problem to a network with deterministic delay for each link.

First, we showed that the value of the solution to the relaxed problem, C_{det} , is very close to the value of the solution to the original problem, C_{T^3} . In fact, as $C_{T^3} \rightarrow \infty$, $\frac{C_{\text{det}}}{C_{T^3}} \rightarrow 1$.

Furthermore, the numerical results indicate that for networks with limited number of clients, the gap between C_{T^3} and C_{det} is very small. Second, we proposed a simple polynomial-time algorithm with additive performance guarantee of N for approximating the relaxed problem. This approximation performs well as the number of Access points is for most cases between 2-4. We also extended the formulation to allow time-varying channels and real-time traffic, and showed the same results for these extensions. Furthermore, for the problem of maximizing weighted total timely throughput we came up with similar results on the bounds on the gap between the original problem and its relaxed version. The future works will investigate more general network models, and application of our proposed deterministic relaxation to other frameworks .

REFERENCES

- [1] "Cisco Visual Networking Index: Forecast and Methodology, 2010-2015," *available at www.cisco.com*, June 2011.
- [2] I-H. Hou, V. Borkar, and P.R. Kumar. A theory of QoS for wireless, "A theory of QoS for wireless," *In Proc. of IEEE INFOCOM*, 2009.
- [3] I-H. Hou, V. Borkar, and P.R. Kumar, "Admission control and scheduling for QoS guarantees for variable-bit-rate applications on wireless channels," *In Proc. of ACM MobiHoc*, 2009.
- [4] I-H. Hou and P.R. Kumar, "Scheduling heterogeneous real-time traffic over fading wireless channels," *In Proc. of IEEE INFOCOM*, 2010.
- [5] L. Tassiulas and A. Ephremides, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Trans. on Information Theory*, Vol. 39, March 1993.
- [6] M. J. Neely, "Delay Analysis for Max Weight Opportunistic Scheduling in Wireless Systems," *IEEE Trans. on Automatic Control*, September 2009.
- [7] A. Dua, C.W. Chan, N. Bambos, and J. Apostolopoulos, "Channel, deadline, and distortion (CD^2) aware scheduling for video streams over wireless," *IEEE Trans. on Wireless Communications*, Vol. 9, No. 3, March 2010.
- [8] M. Agarwal and A. Puri, "Base station scheduling of requests with fixed deadlines," *In Proc. of IEEE INFOCOM*, 487 - 496 Vol.2, 2002.
- [9] M. J. Neely, "Dynamic optimization and learning for renewal systems," *In Proc. of ASILOMAR conference on signals, systems, and computers*, November 2010.
- [10] S. Shakkottai and R. Srikant, "Scheduling real-time traffic with deadlines over a wireless channel," *Wireless Networks*, Vol. 8 Issue 1, January 2002.
- [11] I-H. Hou, A. Truong, S. Chakraborty, and P.R. Kumar, "Optimality of periodwise static priority policies in real-time communications," *To appear in Proc. of CDC*, 2011.
- [12] D.D. Yao, "Dynamic scheduling via polymatroid optimization, performance evaluation of complex systems: techniques and tools, Performance," *Springer-Verlag*, 2002.
- [13] D. B. Shmoys and E. Tardos, "An approximation algorithm for the generalized assignment problem," *Mathematical Programming*, 62:461474, 1993.
- [14] C. Chekuri and S. Khanna, "A PTAS for the Multiple Knapsack Problem," *SIAM Journal on Computing*, 2005.

- [15] L. K. Fleischer, M. X. Goemans, V. S. Mirrokni and M. Sviridenko, “(Almost) Tight Approximation Algorithms for Maximizing General Assignment Problems,” *Symposium on Discrete Algorithms (SODA)*, 2006.
- [16] M.A. Trick, “A Linear relaxation heuristic for the generalized assignment problem,” *Naval Research Logistics*, 1992.
- [17] J.F. Benders and J.A.E.E. van Nunen, “A property of assignment type mixed integer linear programming problems,” *O.R. Letters*, 2, 47-52, 1982.
- [18] K Jain, “A factor 2 approximation algorithm for the generalized Steiner network problem,” *Combinatorica*, Springer, 2001.
- [19] P. Stanica, “Good Lower and Upper Bounds On Binomial Coefficients,” *Journal of Inequalities in Pure and Applied Mathematics*, 2001.

APPENDIX A

PROOF OF THE TIGHTNESS OF THE BOUNDS IN THEOREM 1

We first prove that the upper and lower bounds given in (6) are tight. More specifically, we show that there exist N, M , and some channel success probabilities for which C_{T^3} gets arbitrarily close to $C_{\det} + N$. In addition, there exist N, M , and some channel success probabilities for which $O(|C_{T^3} - C_{\det}|) = O(\sqrt{NC_{\det}})$. Then, we prove Theorem 2.

A. Proof of the tightness of the upper bound

We show that for any given N and $0 < \epsilon < N$ there exist M, τ , and channel success probabilities such that $C_{T^3} - C_{\det} = N - \epsilon$. We set $M = N\tau$, and we choose C_{\det} such that $C_{\det} < M - N$ and $\frac{C_{\det}}{N} \in \mathbb{N}$. Further, for the channel between \mathbf{AP}_i and \mathbf{Rx}_j we set the channel success probability $p_{ij} = \frac{C_{\det} + N - \epsilon}{N\tau}$, where $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, M$. Therefore, according to symmetry, both the optimal assignment which results in C_{T^3} and the optimal assignment for the relaxed problem which results in C_{\det} assign τ packets to each \mathbf{AP} . Furthermore, without loss of generality we can assume that for \mathbf{AP}_i packets of clients $j = 1 + (i-1)\tau, \dots, i\tau$ are assigned to \mathbf{AP}_i . It is easy to check that the following inequalities hold for any \mathbf{AP}_i , $i = 1, 2, \dots, N$:

$$\sum_{j=1+(i-1)\tau}^{1+(i-1)\tau+C_{\det}/N} \frac{1}{p_{ij}} = \left(\frac{C_{\det}}{N}\right) \left(\frac{N\tau}{C_{\det} + N - \epsilon}\right) < \tau < \sum_{j=1+(i-1)\tau}^{1+(i-1)\tau+C_{\det}/N+1} \frac{1}{p_{ij}} \quad (49)$$

Therefore, the maximum number of packets that can be packed in the relaxed problem is C_{\det} . Now, we calculate the expected number of packet deliveries: For any \mathbf{AP}_i the expected number of successful deliveries during one interval is $\tau \left(\frac{C_{\det} + N - \epsilon}{N\tau}\right) = \frac{C_{\det} + N - \epsilon}{N}$. Therefore, we have $C_{T^3} = N \left(\frac{C_{\det} + N - \epsilon}{N}\right) = C_{\det} + N - \epsilon$. Hence, $C_{T^3} - C_{\det} = N - \epsilon$.

B. Proof of the tightness of the order of the lower bound

We show that there exists a wireless network realization for which $O(|C_{T^3} - C_{\text{det}}|) = O(\sqrt{NC_{\text{det}}})$. More specifically, for a given N we show that there exist a positive constant k along with M, τ , such that $C_{\text{det}} - C_{T^3} > k\sqrt{NC_{\text{det}}}$. We choose C_{det} such that $\frac{C_{\text{det}}}{N} \in \mathbb{N}$, and we set $M = C_{\text{det}}$. In addition, we set the channel success probability $p_{ij} = p = \frac{C_{\text{det}}}{N\tau} < 1$ for some $\tau \in \mathbb{N}$, where $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, M$.

Therefore, both the optimal assignment which results in C_{T^3} and the optimal assignment for the relaxed problem which results in C_{det} assign $\frac{C_{\text{det}}}{N}$ packets to each AP. It is easy to check that our chosen C_{det} is actually the solution to the relaxed problem. Now, let Y denote the number of successful deliveries for one of the AP's. Thus, $C_{T^3} = NE[Y]$. Also, let l denote the number of packets that can be packed in a bin corresponding to a certain AP. Therefore, $l = \frac{C_{\text{det}}}{N}$, and $p = \frac{l}{\tau}$. We only need to show that there exists a constant k such that

$$l - E[Y] > k\sqrt{l}. \quad (50)$$

Noting that $l = p\tau$ we have

$$\begin{aligned} l - E[Y] &= p\tau - \left[\sum_{j=1}^l j \binom{\tau}{j} p^j (1-p)^{\tau-j} + l \sum_{j=l+1}^{\tau} \binom{\tau}{j} p^j (1-p)^{\tau-j} \right] \\ &= \sum_{j=1}^{\tau} j \binom{\tau}{j} p^j (1-p)^{\tau-j} - \left[\sum_{j=1}^l j \binom{\tau}{j} p^j (1-p)^{\tau-j} + l \sum_{j=l+1}^{\tau} \binom{\tau}{j} p^j (1-p)^{\tau-j} \right] \\ &= p\tau \sum_{j=l+1}^{\tau} \binom{\tau-1}{j-1} p^{j-1} (1-p)^{\tau-j} - l \sum_{j=l+1}^{\tau} \binom{\tau}{j} p^j (1-p)^{\tau-j} \\ &= l \left[\sum_{j=l+1}^{\tau} \binom{\tau-1}{j-1} p^{j-1} (1-p)^{\tau-j} - \sum_{j=l+1}^{\tau} \left(\binom{\tau-1}{j} + \binom{\tau-1}{j-1} \right) p^j (1-p)^{\tau-j} \right] \\ &= l \left[\sum_{j=l+1}^{\tau} \binom{\tau-1}{j-1} p^{j-1} (1-p)^{\tau-j+1} - \sum_{j=l+1}^{\tau-1} \binom{\tau-1}{j} p^j (1-p)^{\tau-j} \right] \\ &= l \binom{\tau-1}{l} p^l (1-p)^{\tau-l}. \end{aligned}$$

Now note that $\binom{\tau-1}{l} = \frac{(\tau-1)!}{l!(\tau-1-l)!} = \frac{\tau-l}{\tau} \binom{\tau}{l} = (1-p) \binom{\tau}{l}$. Therefore,

$$l \binom{\tau-1}{l} p^l (1-p)^{\tau-l} = T \binom{\tau}{l} p^{l+1} (1-p)^{\tau-l+1}.$$

By Theorem 2.6 of [19] we know that for positive integers m, n, q , with $m > q \geq 1$ and $n \geq 1$

$$\binom{mn}{qn} > \frac{1}{\sqrt{2\pi}} e^{\frac{1}{12n}(\frac{1}{m} - \frac{1}{q} - \frac{1}{m-q})} n^{-\frac{1}{2}} \frac{m^{mn+\frac{1}{2}}}{(m-q)^{(m-q)n+\frac{1}{2}} q^{qn+\frac{1}{2}}}. \quad (51)$$

Substituting n by 1, m by τ , and q by l we get:

$$\begin{aligned} \binom{\tau}{l} &> \frac{1}{\sqrt{2\pi}} \frac{\tau^{\tau+\frac{1}{2}}}{(\tau-l)^{(\tau-l)+\frac{1}{2}} l^{\frac{1}{2}}} e^{\frac{1}{12}(\frac{1}{\tau} - \frac{1}{l} - \frac{1}{\tau-l})} \\ &= \frac{1}{\sqrt{2\pi}} \frac{\tau^{\tau+\frac{1}{2}}}{(\tau(1-p))^{\tau-l+\frac{1}{2}} (p\tau)^{l+\frac{1}{2}}} e^{\frac{1}{12}(\frac{1}{\tau} - \frac{1}{l} - \frac{1}{\tau-l})} \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\tau p(1-p)}} \frac{1}{p^l(1-p)^{\tau-l}} e^{\frac{1}{12}(\frac{1}{\tau} - \frac{1}{l} - \frac{1}{\tau-l})}. \end{aligned}$$

However, $\frac{1}{\tau} - \frac{1}{l} - \frac{1}{\tau-l} > -2$. Therefore,

$$\binom{\tau}{l} > \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\tau p(1-p)}} \frac{1}{p^l(1-p)^{\tau-l}} e^{-\frac{1}{6}}.$$

Hence, we get

$$\begin{aligned} l - E[Y] &= l \binom{\tau-1}{l} p^l (1-p)^{\tau-l} = \tau \binom{\tau}{l} p^{l+1} (1-p)^{\tau-l+1} \\ &> \tau \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\tau p(1-p)}} \frac{1}{p^l(1-p)^{\tau-l}} e^{-\frac{1}{6}} p^{l+1} (1-p)^{\tau-l+1} \\ &> e^{-\frac{1}{6}} \sqrt{\frac{1-p}{2\pi}} \sqrt{l}. \end{aligned}$$

Thus, by setting $k = e^{-\frac{1}{6}} \sqrt{\frac{1-p}{2\pi}}$ the proof will be complete.

APPENDIX B

PROOF OF LEMMA 1

Given that clients are assigned to AP's according to some partition, the problem of finding the optimal scheduling policy is decomposed into finding the optimal scheduling policy for each of the AP's; since AP's are operating at different frequency bands, and have different packets to transmit. In [11] it has been shown that the timely throughput region for the clients assigned to an AP is a polymatroid. Thus, maximum of the summation of timely throughputs for the clients assigned to an AP is achieved at one of the extreme points. Furthermore, it has been shown in [12] that each extreme point of this polymatroid is achieved by a static scheduling policy.

Therefore, there is a static scheduling policy which achieves optimality. Note that when a static scheduling policy is implemented since the channel success probabilities are i.i.d. over intervals, by LLN the total timely throughput is equal to the expected number of packet deliveries during one interval. Therefore, the optimal scheduling policy is the static scheduling policy which results in the maximum expected number of delivered packets for one interval.

We prove that for a given assignment $\vec{\Pi} = [\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_N]$ the optimal ordering of the packet of clients assigned to AP_i is according to the order of p_{ij} , $j \in \mathcal{I}_i$. To prove this, it is sufficient to prove that for any given order if we swap the order of two adjacent clients in such a way that the client with the higher corresponding p_{ij} is prioritized higher, then the expected number of deliveries will be no less than before swapping.

Lemma 5: Let $\tau \in \mathbb{N}$ and G_1, G_2, \dots, G_q be independent geometric random variables with parameters p_1, p_2, \dots, p_q , respectively. Suppose that $p_d < p_{d+1}$ for some $d \in \{1, 2, \dots, q-1\}$. In addition, let G'_1, G'_2, \dots, G'_q be independent geometric random variables (and independent of G_i 's) with parameters $p_1, p_2, \dots, p_{d-1}, p_{d+1}, p_d, p_{d+2}, \dots, p_q$, respectively. Then,

$$\sum_{i=1}^q \Pr\left(\sum_{j=1}^i G_j \leq \tau\right) \leq \sum_{i=1}^q \Pr\left(\sum_{j=1}^i G'_j \leq \tau\right). \quad (52)$$

Proof: We have

$$\begin{aligned} \sum_{i=1}^q \Pr\left(\sum_{j=1}^i G_j \leq \tau\right) &= \sum_{i=1}^{d-1} \Pr\left(\sum_{j=1}^i G_j \leq \tau\right) + \Pr\left(\sum_{j=1}^d G_j \leq \tau\right) + \sum_{i=d+1}^q \Pr\left(\sum_{j=1}^i G_j \leq \tau\right) \\ &= \sum_{i=1}^{d-1} \Pr\left(\sum_{j=1}^i G'_j \leq \tau\right) + \Pr\left(G_d + \sum_{j=1}^{d-1} G'_j \leq \tau\right) + \sum_{i=d+1}^q \Pr\left(\sum_{j=1}^i G'_j \leq \tau\right) \\ &\stackrel{(a)}{\leq} \sum_{i=1}^{d-1} \Pr\left(\sum_{j=1}^i G'_j \leq \tau\right) + \Pr\left(G'_d + \sum_{j=1}^{d-1} G'_j \leq \tau\right) + \sum_{i=d+1}^q \Pr\left(\sum_{j=1}^i G'_j \leq \tau\right) \\ &= \sum_{i=1}^q \Pr\left(\sum_{j=1}^i G'_j \leq \tau\right), \end{aligned}$$

where (a) follows from the fact that success probability of G_d , which is p_d , is less than success probability of G'_d , which is p_{d+1} . ■

According to Lemma 5, the ordering which results in the maximum expected number of delivered packets is the one according to clients' channel success probabilities. So, when a specific partition is chosen, the optimal scheduling policy is the corresponding greedy static scheduling policy.

There are N^M different possible assignments for an interval, each when used together with its corresponding optimal ordering results in a certain expected number of delivered packets. Consequently, the optimal solution to the problem of which assignment (partition) to choose for each interval is simply to choose the assignment which results in the highest expected number of delivered packets, and to repeat it for all intervals.

APPENDIX C

PROOF OF LEMMA 2

For $l = 0$, we have $p_i < \frac{1}{\tau}$, for $i = 1, 2, \dots, q$. Therefore, $E[Y]$ in this case is less than that of the case in which $p_1 = p_2 = \dots = p_q = \frac{1}{\tau}$. On the other hand, for $p_1 = p_2 = \dots = p_q = \frac{1}{\tau}$ $E[Y] \leq \tau * \frac{1}{\tau} = 1$. Hence, the statement is true for $l = 0$. Now, suppose that $l > 0$. We know that $l = \max \hat{l} \text{ s.t. } \sum_{i=1}^{\hat{l}} 1/p_i \leq \tau$. Therefore, we have

$$\sum_{i=1}^l \frac{1}{p_i} \leq \tau < \sum_{i=1}^{l+1} \frac{1}{p_i}. \quad (53)$$

We will show that $E[Y]$ can be at most $l + 1$. Without loss of generality we can omit p_i 's that are equal to zero; because by omitting them neither of $E[Y]$ nor l change, and $E[Y] - l$ would remain the same. So, we suppose that $1 \geq p_1 \geq p_2 \geq \dots \geq p_q > 0$. It is sufficient to prove the lemma for the case of $q = \tau$; because if we have less than τ geometric random variables, $E[Y]$ will be less. On the other hand, we do not need to consider the case $q > \tau$; since for $i > \tau$, $\Pr(\sum_{j=1}^i G_j \leq \tau) = 0$. Therefore, we suppose that $q = \tau$.

Let $X_l = \sum_{i=1}^l G_i$, where $G_i = \text{Geom}(p_i)$. By this notation we have:

$$\Pr(X_l > \tau) = \sum_{i=0}^{l-1} \Pr(Y = i). \quad (54)$$

Now we write down the expression for $E[Y]$:

$$E[Y] = \sum_{i=0}^{l-1} i \Pr(Y = i) + \sum_{t=l}^{\tau} \Pr(X_l = t) \left(\sum_{i=l}^{\tau} i \Pr(Y = i | X_l = t) \right). \quad (55)$$

Since $1 \geq p_l \geq p_{l+1} \geq \dots \geq p_{\tau} > 0$, $E[Y]$ is less than the case where $p_l = p_{l+1} = \dots = p_{\tau}$, although l remains the same. So it is sufficient to prove Theorem 1 for the case where $p_l = p_{l+1} = \dots = p_{\tau}$. For $t \leq \tau$ if we set $p_l = p_{l+1} = \dots = p_{\tau}$ we have

$$\sum_{i=l}^{\tau} i \Pr(Y = i | X_l = t) = E[Y | X_l = t] = l + (\tau - t)p_l. \quad (56)$$

Therefore, by using (55) and (56) we have

$$\begin{aligned}
E[Y] &= \sum_{i=0}^{l-1} i \Pr(Y = i) + \sum_{t=l}^{\tau} (l + p_l(\tau - t)) \Pr(X_l = t) \\
&= \sum_{i=0}^{l-1} i \Pr(Y = i) + (l + p_l\tau)(1 - \Pr(X_l > \tau)) - p_l \left[\sum_{t=l}^{\infty} t \Pr(X_l = t) - \sum_{t=\tau+1}^{\infty} t \Pr(X_l = t) \right] \\
&= \sum_{i=0}^{l-1} i \Pr(Y = i) + (l + p_l\tau) - (l + p_l\tau) \sum_{i=0}^{l-1} \Pr(Y = i) - p_l \sum_{i=1}^l \frac{1}{p_i} + p_l \sum_{t=\tau+1}^{\infty} t \Pr(X_l = t) \\
&= \sum_{i=0}^{l-1} (i - l - p_l\tau) \Pr(Y = i) + (l + p_l(\tau - \sum_{i=1}^l \frac{1}{p_i})) + p_l \sum_{t=\tau+1}^{\infty} t \Pr(X_l = t) \\
&\stackrel{(a)}{<} \sum_{i=0}^{l-1} (i - l - p_l\tau) \Pr(Y = i) + l + 1 + p_l \sum_{t=\tau+1}^{\infty} t \Pr(X_l = t), \tag{57}
\end{aligned}$$

where the last inequality (a) follows from $\tau < \sum_{i=1}^{l+1} \frac{1}{p_i}$ and the assumption that $p_{l+1} = p_l$. Now, we only need to rewrite $p_l \sum_{t=\tau+1}^{\infty} t \Pr(X_l = t)$ in terms of Y . For $t > \tau$ we have

$$\Pr(X_l = t) = \sum_{i=0}^{l-1} \Pr(X_l = t | Y = i) \Pr(Y = i). \tag{58}$$

Therefore,

$$\begin{aligned}
\sum_{t=\tau+1}^{\infty} t \Pr(X_l = t) &= \sum_{t=\tau+1}^{\infty} t \left(\sum_{i=0}^{l-1} \Pr(X_l = t | Y = i) \Pr(Y = i) \right) \\
&= \sum_{i=0}^{l-1} \Pr(Y = i) \left(\sum_{t=\tau+1}^{\infty} t \Pr(X_l = t | Y = i) \right). \tag{59}
\end{aligned}$$

But due to memory less property of geometric distribution, we know that

$$\begin{aligned}
\sum_{t=\tau+1}^{\infty} (t - \tau) \Pr(X_l = t | Y = i) &= \sum_{t=\tau+1}^{\infty} (t - \tau) \Pr\left(\sum_{j=i+1}^l G_j = t - \tau\right) \\
&= \sum_{t=1}^{\infty} t \Pr\left(\sum_{j=i+1}^l G_j = t\right) = \sum_{j=i+1}^l \frac{1}{p_j}, \quad \forall i \leq l-1. \tag{60}
\end{aligned}$$

Therefore,

$$\sum_{t=\tau+1}^{\infty} t \Pr(X_l = t | Y = i) = \tau + \sum_{j=i+1}^l \frac{1}{p_j}. \tag{61}$$

Hence,

$$p_l \sum_{t=\tau+1}^{\infty} t \Pr(X_l = t) = \sum_{i=0}^{l-1} \Pr(Y = i) \left(p_l \tau + \sum_{j=i+1}^l \frac{p_l}{p_j} \right). \tag{62}$$

Substituting (62) into (57) we get

$$\begin{aligned}
E[Y] &< \sum_{i=0}^{l-1} (i - l - p_l \tau) \Pr(Y = i) + l + 1 + \sum_{i=0}^{l-1} \Pr(Y = i) (p_l \tau + \sum_{j=i+1}^l \frac{p_l}{p_j}) \\
&= l + 1 + \sum_{i=0}^{l-1} \Pr(Y = i) (i - l - p_l \tau + p_l \tau + \sum_{j=i+1}^l \frac{p_l}{p_j}) \leq l + 1,
\end{aligned} \tag{63}$$

where the last inequality follows from the fact that $\forall j \in \{i+1, \dots, l\} \quad p_l \leq p_j$.

APPENDIX D

PROOF OF LEMMA 3

We will show that $E[Y]$ is more than $l - 2\sqrt{l + \frac{1}{4}}$. It is sufficient to prove Lemma 3 for the case of $q = l$; because for $q > l$, $E[Y]$ would only increase. On the other hand, q cannot be less than l according to the assumption $l = \max \hat{l} \quad s.t. \quad \sum_{i=1}^{\hat{l}} 1/p_i \leq \tau$. Therefore, from now on we suppose $q = l$. By our notation we have

$$\Pr(\sum_{j=1}^i G_j > \tau) = \sum_{j=0}^{i-1} \Pr(Y = j), \quad i = 1, 2, \dots, l. \tag{64}$$

We now bound $l - E[Y]$ from above.

$$\begin{aligned}
l - E[Y] &= l - \sum_{i=1}^l \Pr(Y \geq i) = \sum_{i=1}^l (1 - \Pr(Y \geq i)) = \sum_{i=1}^l \Pr(Y < i) \\
&\stackrel{(a)}{=} \sum_{i=1}^l \Pr(\sum_{j=1}^i G_j > \tau) \stackrel{(b)}{\leq} \sum_{i=1}^l \Pr(\sum_{j=1}^i G_j > \sum_{j=1}^l \frac{1}{p_j}) \\
&\leq 1 + \sum_{i=1}^{l-1} \Pr(|\sum_{j=1}^i (G_j - \frac{1}{p_j})| > \sum_{j=i+1}^l \frac{1}{p_j}),
\end{aligned}$$

where (a) follows from (64); and (b) follows from $\sum_{i=1}^l 1/p_i \leq \tau$. Now, by using Chebyshev's inequality

$$\Pr(|\sum_{j=1}^i (G_j - \frac{1}{p_j})| > \sum_{j=i+1}^l \frac{1}{p_j}) \leq \min(1, \frac{\text{var}(\sum_{j=1}^i G_j)}{(\sum_{j=i+1}^l \frac{1}{p_j})^2}), \quad \forall i \in \mathbb{N}, 1 \leq i \leq l-1, \tag{65}$$

where $\text{var}(\sum_{j=1}^i G_j)$ is the variance of the random variable $\sum_{j=1}^i G_j$. Therefore,

$$l - E[Y] \leq 1 + \sum_{i=1}^{l-1} \min(1, \frac{\text{var}(\sum_{j=1}^i G_j)}{(\sum_{j=i+1}^l \frac{1}{p_j})^2}).$$

Due to independence of G_i 's, we have

$$\text{var}\left(\sum_{j=1}^i G_j\right) = \sum_{j=1}^i \text{var}(G_j) = \sum_{j=1}^i \frac{1-p_j}{p_j^2} < \sum_{j=1}^i \frac{1}{p_j^2}, \quad i = 1, 2, \dots, l. \quad (66)$$

Therefore,

$$l - E[Y] \leq 1 + \sum_{i=1}^{l-1} \min\left(1, \frac{\sum_{j=1}^i \frac{1}{p_j^2}}{(\sum_{j=i+1}^l \frac{1}{p_j})^2}\right). \quad (67)$$

But since $p_1 \geq p_2 \geq \dots \geq p_l$, we have $\sum_{j=1}^i \frac{1}{p_j^2} \leq \frac{i}{p_i^2}$ and $(\sum_{j=i+1}^l \frac{1}{p_j})^2 \geq (\frac{l-i}{p_i})^2$. Therefore,

$$l - E[Y] \leq 1 + \sum_{i=1}^{l-1} \min\left(1, \frac{\frac{i}{p_i^2}}{(\frac{l-i}{p_i})^2}\right) = 1 + \sum_{i=1}^{l-1} \min\left(1, \frac{i}{(l-i)^2}\right). \quad (68)$$

Hence, by (68) and applying Lemma 6, which is stated and proved as following, the proof of Lemma 3 will be complete.

Lemma 6: Assume $l \in \mathbb{N}$, and $l > 1$. Then,

$$1 + \sum_{i=1}^{l-1} \min\left(1, \frac{i}{(l-i)^2}\right) < 2\sqrt{l + \frac{1}{4}}. \quad (69)$$

Proof: For $l < 18$ the statement of the Lemma can be verified numerically. Therefore, suppose that $l \geq 18$. Let $f(i) \triangleq \frac{i}{(l-i)^2}$, for $i \in \mathbb{N}, 1 \leq i \leq l-1$; and consider the following three observations regarding the function $f(\cdot)$:

- 1) $f(i)$ increases as i increases for $i \in \mathbb{N}, 1 \leq i \leq l-1$.
- 2) $f(1) = \frac{1}{(l-1)^2} < 1$.
- 3) $f(l-1) = \frac{l-1}{1} > 1$

Therefore, $\exists m \in \mathbb{N}, 1 \leq m < l-1$ such that

$$\frac{m}{(l-m)^2} \leq 1 < \frac{m+1}{(l-(m+1))^2}. \quad (70)$$

Note that m cannot be equal to $l-1$, because $\frac{l-1}{(l-(l-1))^2} > 1$. Rewriting the inequalities in (70) we get

$$l - \sqrt{l + \frac{1}{4}} - \frac{1}{2} < m \leq l - \sqrt{l + \frac{1}{4}} + \frac{1}{2}. \quad (71)$$

Now, we rewrite $1 + \sum_{i=1}^{l-1} \min(1, \frac{i}{(l-i)^2})$:

$$1 + \sum_{i=1}^{l-1} \min\left(1, \frac{i}{(l-i)^2}\right) = 1 + \sum_{i=1}^m \min\left(1, \frac{i}{(l-i)^2}\right) + \sum_{i=m+1}^{l-1} \min\left(1, \frac{i}{(l-i)^2}\right). \quad (72)$$

But from (70) and the fact that $f(i) = \frac{i}{(l-i)^2}$ increases by increase of i , we have

$$\begin{aligned}
1 + \sum_{i=1}^{l-1} \min(1, \frac{i}{(l-i)^2}) &= 1 + \sum_{i=1}^m \frac{i}{(l-i)^2} + (l-1-m) = l-m + \sum_{j=l-m}^{l-1} \frac{l-j}{j^2} \\
&< l-m + \sum_{j=l-m}^{l-1} \frac{l-j}{j(j-1)} = l-m + \sum_{j=l-m}^{l-1} (\frac{l-j}{j-1} - \frac{l-j}{j}) = l-m + \frac{m}{l-m-1} - \sum_{j=l-m}^{l-1} \frac{1}{j} \\
&\stackrel{(a)}{<} l-m + \frac{m}{l-m-1} - \frac{m}{l-\frac{m+1}{2}} = l-m + \frac{m(m+1)}{(l-m-1)(l-m+l-1)} \\
&\stackrel{(b)}{<} (\sqrt{l+\frac{1}{4}} + \frac{1}{2}) + \frac{(l-\sqrt{l+\frac{1}{4}}+\frac{1}{2})(l-\sqrt{l+\frac{1}{4}}+\frac{3}{2})}{(\sqrt{l+\frac{1}{4}}-\frac{3}{2})(l+\sqrt{l+\frac{1}{4}}-\frac{3}{2})} \\
&= (\sqrt{l+\frac{1}{4}} + \frac{1}{2}) + (\sqrt{l+\frac{1}{4}} - \frac{3}{2} + \frac{5l-9\sqrt{l+\frac{1}{4}}+\frac{11}{2}}{(\sqrt{l+\frac{1}{4}}-\frac{3}{2})(l+\sqrt{l+\frac{1}{4}}-\frac{3}{2})}) \\
&= 2\sqrt{l+\frac{1}{4}} + \frac{-l\sqrt{l+\frac{1}{4}}+\frac{11}{2}l-6\sqrt{l+\frac{1}{4}}+3}{(\sqrt{l+\frac{1}{4}}-\frac{3}{2})(l+\sqrt{l+\frac{1}{4}}-\frac{3}{2})}, \tag{73}
\end{aligned}$$

where (a) follows from the Cauchy-Schwarz inequality; and (b) follows from (71). For $l \geq 18$ the term $\frac{-l\sqrt{l+\frac{1}{4}}+\frac{11}{2}l-6\sqrt{l+\frac{1}{4}}+3}{(\sqrt{l+\frac{1}{4}}-\frac{3}{2})(l+\sqrt{l+\frac{1}{4}}-\frac{3}{2})}$ in (73) is less than zero. Therefore, the statement of Lemma 6 is true for all $l > 1, l \in \mathbb{N}$. ■

APPENDIX E

PROOF OF COROLLARY 1

Let $\vec{\Pi}^*$ denote the partition (assignment) chosen by the optimal greedy static scheduling policy $\eta_{\text{g-static}}^*$. Therefore, we have $\|\vec{R}(\eta_{\text{g-static}}^*)\|_1 = C_{\text{T}^3}$. Furthermore, consider an assignment, denoted by $\vec{\Pi}_{\text{det}}$, which maximizes the objective function in (5). Let $\eta_{\text{g-static}}^{\text{det}}$ denote the greedy static scheduling policy which corresponds to $\vec{\Pi}_{\text{det}}$. Further, let $\|\vec{R}_{\text{det}}(\eta_{\text{static}})\|_1$ designate the maximum number of objects that can be packed in the RP in (5) when a static scheduling policy η_{static} is implemented. Therefore, $\|\vec{R}_{\text{det}}(\eta_{\text{g-static}}^{\text{det}})\|_1 = C_{\text{det}}$, since $\|\vec{R}_{\text{det}}(\eta_{\text{g-static}}^{\text{det}})\|_1$ is the value of the objective function in (5) when the assignment is dictated by $\eta_{\text{g-static}}^{\text{det}}$. The right part of the inequality in Corollary 1 in (7) is trivial since C_{T^3} is the optimal T^3 achievable under any scheduling policy. So we only need to prove the left part of the inequality in (7). Using a similar argument as the

one in part B of Section IV, and by applying Cauchy-Schwarz inequality, we get

$$\|\vec{R}(\eta_{\text{g-static}}^{\text{det}})\|_1 \geq \|\vec{R}_{\text{det}}(\eta_{\text{g-static}}^{\text{det}})\|_1 - 2\sqrt{N(\|\vec{R}_{\text{det}}(\eta_{\text{g-static}}^{\text{det}})\|_1 + \frac{N}{4})}. \quad (74)$$

Now consider the function $g(\cdot)$ defined as follows: $g(x) \triangleq x - 2\sqrt{N(x + \frac{N}{4})}$, $x \in \mathbb{R}$.

So, $g(x)$ is strictly increasing for $x > \frac{3N}{4}$. On the other hand, we know that

$$C_{\text{det}} = \|\vec{R}_{\text{det}}(\eta_{\text{g-static}}^{\text{det}})\|_1 \geq \|\vec{R}_{\text{det}}(\eta_{\text{g-static}}^*)\|_1 \geq \|\vec{R}(\eta_{\text{g-static}}^*)\|_1 - N = C_{\text{T}^3} - N, \quad (75)$$

where the right inequality follows from Theorem 1. By $g(x)$ being an increasing function of x and (75) we get

$$\|\vec{R}_{\text{det}}(\eta_{\text{g-static}}^{\text{det}})\|_1 - 2\sqrt{N(\|\vec{R}_{\text{det}}(\eta_{\text{g-static}}^{\text{det}})\|_1 + \frac{N}{4})} \geq C_{\text{T}^3} - N - 2\sqrt{N(C_{\text{T}^3} - \frac{3N}{4})}. \quad (76)$$

Hence, by (74) and (76) we get

$$\|\vec{R}(\eta_{\text{g-static}}^{\text{det}})\|_1 \geq C_{\text{T}^3} - N - 2\sqrt{N(C_{\text{T}^3} - \frac{3N}{4})}. \quad (77)$$

APPENDIX F

PROOF OF THEOREM 4

By the same argument as the one in proof of Lemma 1, $C_{w-\text{T}^3}$ can be achieved by a static scheduling policy in which the order of a client Rx_j is according to the value of $\omega_j p_{ij}$ when assigned to AP_i . Therefore, by LLN, in order to achieve $C_{w-\text{T}^3}$, it is sufficient to find the assignment which has the highest expected weighted delivery for one interval. We start by proving the right part of the inequality in (48).

First, we prove that for a given assignment $\vec{\Pi} = [\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_N]$ the optimal ordering of the packet of clients assigned to AP_i is according to the order of $\omega_j p_{ij}$, $j \in \mathcal{I}_i$. To do so, it is sufficient to prove that for any given order of the clients if we swap the order of two adjacent clients in such a way that the client with higher $\omega_j p_j$ is prioritized higher, then the expected weighted delivery will be no less than before swapping.

Lemma 7: Let $\tau, q \in \mathbb{N}$, and $\omega_1, \omega_2, \dots, \omega_q \in \mathbb{R}$. Also, for some $d \in \{1, 2, \dots, q-1\}$, let $\omega'_i = \omega_i$, for $1 \leq i < d$ and $d+1 < i \leq q$; and $\omega'_d = \omega_{d+1}$, $\omega'_{d+1} = \omega_d$. Further, let G_1, G_2, \dots, G_q be independent geometric random variables with parameters p_1, p_2, \dots, p_q , respectively. Suppose

that $\omega_d p_d < \omega_{d+1} p_{d+1}$. In addition, let G'_1, G'_2, \dots, G'_q be independent geometric random variables, independent of G_i 's, with parameters $p_1, p_2, \dots, p_{d-1}, p_{d+1}, p_d, p_{d+2}, \dots, p_q$, respectively.

Then,

$$\sum_{i=1}^q \omega_i \Pr(\sum_{j=1}^i G_j \leq \tau) \leq \sum_{i=1}^q \omega'_i \Pr(\sum_{j=1}^i G'_j \leq \tau). \quad (78)$$

Proof: Let $A = \sum_{i=1}^q \omega_i \Pr(\sum_{j=1}^i G_j \leq \tau)$, and $B = \sum_{i=1}^q \omega'_i \Pr(\sum_{j=1}^i G'_j \leq \tau)$. Then,

$$\begin{aligned} B - A &= \sum_{i=1}^{d-1} \omega'_i \Pr(\sum_{j=1}^i G'_j \leq \tau) + \omega'_d \Pr(\sum_{j=1}^d G'_j \leq \tau) + \omega'_{d+1} \Pr(\sum_{j=1}^{d+1} G'_j \leq \tau) \\ &\quad + \sum_{i=d+1}^q \omega'_i \Pr(\sum_{j=1}^i G'_j \leq \tau) - \sum_{i=1}^{d-1} \omega_i \Pr(\sum_{j=1}^i G_j \leq \tau) - \omega_d \Pr(\sum_{j=1}^d G_j \leq \tau) \\ &\quad - \omega_{d+1} \Pr(\sum_{j=1}^{d+1} G_j \leq \tau) - \sum_{i=d+1}^q \omega_i \Pr(\sum_{j=1}^i G_j \leq \tau) \\ &= \omega'_d \Pr(\sum_{j=1}^d G'_j \leq \tau) + \omega'_{d+1} \Pr(\sum_{j=1}^{d+1} G'_j \leq \tau) - \omega_d \Pr(\sum_{j=1}^d G_j \leq \tau) \\ &\quad - \omega_{d+1} \Pr(\sum_{j=1}^{d+1} G_j \leq \tau) \\ &= \sum_{t=1}^{\tau} \Pr(\sum_{j=1}^{d-1} G'_j = t) [\omega'_d \Pr(G'_d \leq \tau - t) + \omega'_{d+1} \Pr(G'_d + G'_{d+1} \leq \tau - t)] \\ &\quad - \sum_{t=1}^{\tau} \Pr(\sum_{j=1}^{d-1} G_j = t) [\omega_d \Pr(G_d \leq \tau - t) + \omega_{d+1} \Pr(G_d + G_{d+1} \leq \tau - t)] \\ &= \sum_{t=1}^{\tau} \Pr(\sum_{j=1}^{d-1} G_j = t) [\omega'_d \Pr(G'_d \leq \tau - t) + \omega'_{d+1} \Pr(G'_d + G'_{d+1} \leq \tau - t) \\ &\quad - \omega_d \Pr(G_d \leq \tau - t) - \omega_{d+1} \Pr(G_d + G_{d+1} \leq \tau - t)]. \end{aligned}$$

Therefore, it is sufficient to show that for all $t \in \mathbb{N}$,

$$\omega'_d \Pr(G'_d \leq t) + \omega'_{d+1} \Pr(G'_d + G'_{d+1} \leq t) - \omega_d \Pr(G_d \leq t) - \omega_{d+1} \Pr(G_d + G_{d+1} \leq t) \geq 0.$$

Note that

- $\omega'_d = \omega_{d+1}$, and $\omega'_{d+1} = \omega_d$.
- $\Pr(G'_d \leq t) = 1 - (1 - p_{d+1})^t$, and $\Pr(G_d \leq t) = 1 - (1 - p_d)^t$.
- $\Pr(G_d + G_{d+1} \leq t) = \Pr(G'_d + G'_{d+1} \leq t) = 1 - \frac{p_d(1-p_{d+1})^t - p_{d+1}(1-p_d)^t}{p_d - p_{d+1}}$.

Therefore,

$$\omega'_d \Pr(G'_d \leq t) + \omega'_{d+1} \Pr(G'_d + G'_{d+1} \leq t) - \omega_d \Pr(G_d \leq t) - \omega_{d+1} \Pr(G_d + G_{d+1} \leq t) \quad (79)$$

$$= (\omega_{d+1}p_{d+1} - \omega_d p_d) \left(\frac{(1 - p_{d+1})^t - (1 - p_d)^t}{p_d - p_{d+1}} \right) > 0, \quad t \in \mathbb{N}, \quad (80)$$

where the inequality follows from the assumption that $\omega_{d+1}p_{d+1} - \omega_d p_d > 0$. \blacksquare

A. Proof of $C_{w-\mathsf{T}^3} < C_{w-\text{det}} + N\omega_{\max}$

We follow the same line of proof as in Section IV. Since by the same argument made in the proof of Lemma 1, $C_{w-\mathsf{T}^3}$ can be achieved using a static scheduling policy which uses the optimal ordering of clients according to $\omega_j p_j$'s, it is sufficient to show that for any static scheduling policy $\eta_{\text{wg-static}}$ which uses its corresponding optimal ordering we have

$$w-\mathsf{T}^3(\eta_{\text{wg-static}}) < C_{w-\text{det}} + N\omega_{\max}. \quad (81)$$

Suppose an arbitrary static scheduling policy $\eta_{\text{wg-static}}$ with the corresponding partition $\vec{\Pi}_{\text{wg-static}} = [\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_N]$, which uses the optimal ordering is implemented. By (45) we know that

$$w-\mathsf{T}^3(\eta_{\text{wg-static}}) = \sum_{j=1}^M \omega_j R_j(\eta_{\text{wg-static}}). \quad (82)$$

On the other hand, by (1) we have

$$R_j(\eta_{\text{wg-static}}) = \liminf_{r \rightarrow \infty} \frac{\sum_{k=1}^r N_j(k, \eta_{\text{wg-static}})}{r}, \quad j \in \{1, 2, \dots, M\}. \quad (83)$$

Let Y_i denote the random variable indicating the total number of successful deliveries by \mathbf{AP}_i during one interval, when $\eta_{\text{wg-static}}$ is implemented, $i = 1, 2, \dots, N$. More precisely,

$$Y_i \triangleq \sum_{j \in \mathcal{I}_i} N_j(1, \eta_{\text{wg-static}}), \quad i = 1, 2, \dots, N. \quad (84)$$

Define $q_i \triangleq |\mathcal{I}_i|$. Denote the enumeration of clients assigned to \mathbf{AP}_i by $\{\mathcal{I}_i(1), \mathcal{I}_i(2), \dots, \mathcal{I}_i(q_i)\}$, where the enumeration is according to the optimal ordering for the weighted case. Since a static scheduling policy is implemented and channels are i.i.d over time, by LLN we have

$$R_{\mathcal{I}_i(j)}(\eta_{\text{wg-static}}) = \liminf_{r \rightarrow \infty} \frac{\sum_{k=1}^r N_{\mathcal{I}_i(j)}(k, \eta_{\text{wg-static}})}{r} = \Pr(Y_i \geq j), \quad 1 \leq j \leq q_i, \quad 1 \leq i \leq N.$$

Therefore,

$$\begin{aligned}
 w\text{-T}^3(\eta_{\text{wg-static}}) &= \sum_{i=1}^N \sum_{j \in \mathcal{I}_i} \omega_j R_j(\eta_{\text{wg-static}}) = \sum_{i=1}^N \sum_{j=1}^{q_i} \omega_{\mathcal{I}_i(j)} \Pr(Y_i \geq j) \\
 &= \sum_{i=1}^N \sum_{j=1}^{q_i} \left(\sum_{k=1}^j \omega_{\mathcal{I}_i(k)} \right) \Pr(Y_i = j).
 \end{aligned} \tag{85}$$

Let G_{ij} be a geometric random variable with parameter p_{ij} , $i = 1, 2, \dots, N$, $j = 1, 2, \dots, M$.

Then, it is easy to see that

$$Y_i = \max k \quad s.t. \quad \sum_{j=1}^k G_{i\mathcal{I}_i(j)} \leq \tau, \quad i \in \{1, 2, \dots, N\}, k \leq q_i, \tag{86}$$

since $\eta_{\text{wg-static}}$ persistently sends a packet until it is delivered, or the interval is over. The following lemma relates l_i and ω_j 's to Y_i .

Lemma 8: Let $1 \leq \omega_1, \omega_2, \dots, \omega_q \leq \omega_{\max}$ for some $\omega_{\max} \in \mathbb{R}$. Also, let $\tau \in \mathbb{N}$ and G_1, G_2, \dots, G_q be independent geometric random variables with parameters p_1, p_2, \dots, p_q respectively, such that $\omega_1 p_1 \geq \omega_2 p_2 \geq \dots \geq \omega_q p_q \geq 0$. Also define

$$l \triangleq \max \hat{l} \quad s.t. \quad \sum_{i=1}^{\hat{l}} 1/p_i \leq \tau, \tag{87}$$

and

$$Y \triangleq \max i \quad s.t. \quad \sum_{j=1}^i G_j \leq \tau, \quad i \in \{1, 2, \dots, q\}. \tag{88}$$

Then, we have

$$\sum_{i=1}^q \left(\sum_{j=1}^i \omega_j \right) \Pr(Y = i) < \sum_{j=1}^l \omega_j + \omega_{\max}. \tag{89}$$

Hence, by Lemma 8 we have

$$\sum_{i=1}^N \sum_{j=1}^{q_i} \left(\sum_{k=1}^j \omega_{\mathcal{I}_i(k)} \right) \Pr(Y_i = j) < \sum_{i=1}^N \sum_{j=1}^{l_i} \omega_{\mathcal{I}_i(j)} + N\omega_{\max}. \tag{90}$$

Therefore, by (85) and (90) we have

$$w\text{-T}^3(\eta_{\text{wg-static}}) < \sum_{i=1}^N \sum_{j=1}^{l_i} \omega_{\mathcal{I}_i(j)} + N\omega_{\max}. \tag{91}$$

But, $\sum_{i=1}^N \sum_{j=1}^{l_i} \omega_{\mathcal{I}_i(j)}$ is the value of the objective function in (47) for a feasible solution.

Therefore,

$$\sum_{i=1}^N \sum_{j=1}^{l_i} \omega_{\mathcal{I}_i(j)} \leq C_{w\text{-det}}. \tag{92}$$

Putting (91) and (92) together we get

$$w\text{-}\mathsf{T}^3(\eta_{\text{wg-static}}) < C_{w\text{-det}} + N\omega_{\max}. \quad (93)$$

Hence the proof of the right inequality in Theorem 4 is complete. We now prove Lemma 8.

Proof: Suppose that $l > 0$ (for $l = 0$ the proof is straightforward). We have

$$\sum_{i=1}^l \frac{1}{p_i} \leq \tau < \sum_{i=1}^{l+1} \frac{1}{p_i}. \quad (94)$$

Without loss of generality we can omit p_i 's that are equal to zero and assume $0 < p_1, p_2, \dots, p_q \leq$

1. Furthermore, according to the same argument as in proof of Theorem 1, it is sufficient to prove the lemma for the case of $q = \tau$. Let $X_l = \sum_{i=1}^l G_i$, where $G_i = \text{Geom}(p_i)$. We have

$$\begin{aligned} \sum_{i=1}^{\tau} \left(\sum_{j=1}^i \omega_j \right) \Pr(Y = i) &= \sum_{j=1}^{l-1} \left(\sum_{j=1}^i \omega_j \right) \Pr(Y = i) \\ &+ \sum_{t=1}^{\tau} \Pr(X_l = t) \left(\sum_{i=l}^{\tau} \left(\sum_{j=1}^i \omega_j \right) \Pr(Y = i | X_l = t) \right) \end{aligned}$$

However, since $\omega_{\max} \geq \omega_l p_l \geq \omega_{l+1} p_{l+1} \geq \dots \geq \omega_{\tau} p_{\tau} > 0$, $\sum_{i=1}^{\tau} \left(\sum_{j=1}^i \omega_j \right) \Pr(Y = i)$ is less than the case where $\omega_l p_l = \omega_{l+1} p_{l+1} = \dots = \omega_{\tau} p_{\tau}$. With a similar argument as in the proof of Theorem 1 we get

$$\begin{aligned} \sum_{i=1}^{\tau} \left(\sum_{j=1}^i \omega_j \right) \Pr(Y = i) &\leq \sum_{i=1}^{l-1} \left(\sum_{j=1}^i \omega_j \right) \Pr(Y = i) + \sum_{t=1}^{\tau} \left(\sum_{j=1}^l \omega_j + \omega_l p_l (\tau - t) \right) \Pr(X_l = t) \\ &= \sum_{i=1}^{l-1} \left(\sum_{j=1}^i \omega_j \right) \Pr(Y = i) + \left(\sum_{j=1}^l \omega_j + \omega_l p_l \tau \right) (1 - \Pr(X_l > \tau)) \\ &- \omega_l p_l \left[\sum_{t=1}^{\infty} t \Pr(X_l = t) - \sum_{t=\tau+1}^{\infty} t \Pr(X_l = t) \right] = \sum_{i=1}^{l-1} \left(\sum_{j=1}^i \omega_j \right) \Pr(Y = i) + \left(\sum_{j=1}^l \omega_j + \omega_l p_l \tau \right) \\ &- \left(\sum_{j=1}^l \omega_j + \omega_l p_l \tau \right) \sum_{i=0}^{l-1} \Pr(Y = i) - \omega_l p_l \sum_{i=1}^l \frac{1}{p_i} + \omega_l p_l \sum_{t=\tau+1}^{\infty} t \Pr(X_l = t) \\ &= \sum_{i=0}^{l-1} \left(\sum_{j=1}^i \omega_j - \sum_{j=1}^l \omega_j - \omega_l p_l \tau \right) \Pr(Y = i) + \left(\sum_{j=1}^l \omega_j + \omega_l p_l \left(\tau - \sum_{i=1}^l \frac{1}{p_i} \right) \right) \\ &+ \omega_l p_l \sum_{t=\tau+1}^{\infty} t \Pr(X_l = t) \\ &\stackrel{(a)}{<} \sum_{i=0}^{l-1} \left(\sum_{j=1}^i \omega_j - \sum_{j=1}^l \omega_j - \omega_l p_l \tau \right) \Pr(Y = i) + \left(\sum_{j=1}^l \omega_j + \omega_{l+1} \right) + \omega_l p_l \sum_{t=\tau+1}^{\infty} t \Pr(X_l = t) \end{aligned}$$

$$\begin{aligned}
&\stackrel{(b)}{=} \sum_{i=0}^{l-1} \left(- \sum_{j=i+1}^l \omega_j - \omega_l p_l \tau \right) \Pr(Y=i) + \left(\sum_{j=1}^l \omega_j + \omega_{l+1} \right) + \sum_{i=0}^{l-1} \Pr(Y=i) \left(\omega_l p_l \tau + \sum_{j=i+1}^l \frac{\omega_l p_l}{p_j} \right) \\
&= \left(\sum_{j=1}^l \omega_j + \omega_{l+1} \right) + \sum_{i=0}^{l-1} \left(\sum_{j=i+1}^l \frac{\omega_l p_l}{p_j} - \sum_{j=i+1}^l \omega_j \right) \Pr(Y=i) \stackrel{(c)}{\leq} \sum_{j=1}^l \omega_j + \omega_{l+1} \leq \sum_{j=1}^l \omega_j + \omega_{l+1}.
\end{aligned}$$

where (a) follows from $\tau - \sum_{i=1}^l \frac{1}{p_i} < \frac{1}{p_{l+1}}$ and $\omega_l p_l = \omega_{l+1} p_{l+1}$; (b) follows from (62); and (c) follows from the fact that $\forall j \in \{i+1, \dots, l\} \quad \frac{\omega_l p_l}{p_j} \leq \omega_j$. \blacksquare

B. Proof of $C_{w\text{-det}} - 2\omega_{\max} \sqrt{N(C_{w\text{-det}} + \frac{N}{4})} < C_{w\text{-T}^3}$

The proof of the lower bound is similar to the proof of lower bound in Theorem 1. Consider the assignment proposed by the solution to the problem in (47), where the clients which have not been assigned to any AP for transmission are assigned to AP's arbitrarily. Let $\vec{\Pi}_{\text{wg-static}}^{\text{det}} = [\mathcal{I}_1^{\text{det}}, \mathcal{I}_2^{\text{det}}, \dots, \mathcal{I}_N^{\text{det}}]$ denote the resulting partition, and also let $\eta_{\text{wg-static}}^{\text{det}}$ denote the corresponding static scheduling policy which orders clients based on their channel success probabilities. Therefore, $w\text{-T}^3(\eta_{\text{wg-static}}^{\text{det}}) \leq C_{w\text{-T}^3}$. So, it is sufficient to prove that $C_{w\text{-det}} - 2\sqrt{N(C_{w\text{-det}} + \frac{N}{4})} < w\text{-T}^3(\eta_{\text{wg-static}}^{\text{det}})$. Let W_i^{det} denote the random variable indicating the total weight of successful deliveries by AP_{*i*} during one interval, when $\eta_{\text{wg-static}}^{\text{det}}$ is implemented, $i = 1, 2, \dots, N$. More precisely,

$$W_i \triangleq \sum_{j \in \mathcal{I}_i^{\text{det}}} \omega_j N_j(1, \eta_{\text{wg-static}}), \quad i = 1, 2, \dots, N. \quad (95)$$

Then, by LLN we have

$$w\text{-T}^3(\eta_{\text{wg-static}}^{\text{det}}) = \sum_{i=1}^N E[W_i^{\text{det}}].$$

Therefore, it is sufficient to prove that

$$C_{w\text{-det}} - 2\sqrt{N(C_{w\text{-det}} + \frac{N}{4})} < \sum_{i=1}^N E[W_i^{\text{det}}].$$

Define $q_i = |\mathcal{I}_i^{\text{det}}|$, and denote the enumeration of clients assigned to AP_{*i*} by $\{\mathcal{I}_i^{\text{det}}(1), \mathcal{I}_i^{\text{det}}(2), \dots, \mathcal{I}_i^{\text{det}}(q_i)\}$, where the enumeration is according to the channel success probabilities of different clients in \mathcal{I}_i . Further, let G_{ij} be a geometric random variable with parameter p_{ij} , $i = 1, 2, \dots, N$, $j = 1, 2, \dots, M$. It is easy to see that

$$W_i^{\text{det}} = \max \sum_{j=1}^k \omega_j \quad s.t. \quad \sum_{j=1}^k G_{i\mathcal{I}_i(j)^{\text{det}}} \leq \tau, \quad i \in \{1, 2, \dots, N\}, k \leq q_i, \quad (96)$$

since $\eta_{\text{g-static}}^{\text{det}}$ persistently sends a packet until it is delivered, or the interval is over. Also define

$$l_i^{\text{det}} \triangleq \max_{\hat{l}} \quad \hat{l} \quad \text{s.t.} \quad \sum_{j=1}^{\hat{l}} 1/p_{i\mathcal{I}_i^{\text{det}}(j)} \leq \tau, \quad \hat{l} \leq q_i.$$

Therefore, l_i^{det} is the maximum number of objects that fit into a bin of capacity τ when the channels are relaxed and clients in $\mathcal{I}_i^{\text{det}}$ are assigned to AP_i . By Lemma 9, which is stated and proved later in this appendix, we have

$$\sum_{i=1}^N \sum_{j=1}^{l_i^{\text{det}}} \omega_{\mathcal{I}_i(j)} - 2\omega_{\max} \sum_{i=1}^N \sqrt{l_i^{\text{det}} + \frac{1}{4}} < \sum_{i=1}^N E[W_i^{\text{det}}]. \quad (97)$$

Note that $\sum_{i=1}^N \sum_{j=1}^{l_i^{\text{det}}} \omega_{\mathcal{I}_i(j)} = C_{w\text{-det}}$. According to Cauchy-Schwarz inequality we have

$$\begin{aligned} \omega_{\max} \sum_{i=1}^N \sqrt{l_i^{\text{det}} + \frac{1}{4}} &\leq \omega_{\max} \sqrt{N \left(\sum_{i=1}^N l_i^{\text{det}} + \frac{N}{4} \right)} \\ &\leq \omega_{\max} \sqrt{N \left(\sum_{i=1}^N \sum_{j=1}^{l_i^{\text{det}}} \omega_{\mathcal{I}_i(j)} + \frac{N}{4} \right)} = \omega_{\max} \sqrt{N(C_{w\text{-det}} + \frac{N}{4})}. \end{aligned} \quad (98)$$

Putting (97) and (98) together we get

$$C_{w\text{-det}} - 2\sqrt{N(C_{w\text{-det}} + \frac{N}{4})} < C_{w\text{-T}^3}. \quad (99)$$

Hence, the left inequality of Theorem 4 is proved and the proof of Theorem 4 is complete. We now state and prove Lemma 9.

Lemma 9: Let $1 \leq \omega_1, \omega_2, \dots, \omega_q \leq \omega_{\max}$ for some $\omega_{\max} \in \mathbb{R}$. Also, let $\tau \in \mathbb{N}$ and G_1, G_2, \dots, G_q be independent geometric random variables with parameters p_1, p_2, \dots, p_q respectively, such that $1 \geq p_1 \geq p_2 \geq \dots \geq p_q \geq 0$. Also define

$$l \triangleq \max_{\hat{l}} \quad \hat{l} \quad \text{s.t.} \quad \sum_{i=1}^{\hat{l}} 1/p_i \leq \tau, \quad (100)$$

and

$$Y \triangleq \max_i i \quad \text{s.t.} \quad \sum_{j=1}^i G_j \leq \tau, \quad i \in \{1, 2, \dots, q\}. \quad (101)$$

Then, we have

$$\sum_{j=1}^l \omega_j - 2\omega_{\max} \sqrt{l + \frac{1}{4}} < \sum_{i=1}^q \left(\sum_{j=1}^i \omega_j \right) \Pr(Y = i). \quad (102)$$

Proof: With the same argument as in Theorem 1, it is sufficient to assume $q = l$. The proof is very similar to the proof of lower bound in Theorem 1:

$$\begin{aligned}
& \sum_{i=1}^l \omega_i - \sum_{i=1}^l \left(\sum_{j=1}^i \omega_j \right) \Pr(Y = i) = \sum_{i=1}^l \omega_i - \sum_{i=1}^l \omega_i \left(\sum_{j=i}^l \Pr(Y = j) \right) \\
& = \sum_{i=1}^l \omega_i \Pr(G_1 + G_2 + \dots + G_i > \tau) \\
& \leq \sum_{i=1}^{l-1} \omega_i \Pr\left(\left| \sum_{j=1}^i (G_j - \frac{1}{p_j}) \right| > \sum_{j=i+1}^l \frac{1}{p_j} \right) + \omega_l \stackrel{(a)}{\leq} \sum_{i=1}^{l-1} \omega_i \min\left(1, \frac{\text{var}(\sum_{j=1}^i G_j)}{(\sum_{j=i+1}^l \frac{1}{p_j})^2}\right) + \omega_l \\
& \leq \sum_{i=1}^{l-1} \omega_i \min\left(1, \frac{\sum_{j=1}^i \frac{1}{p_j^2}}{(\sum_{j=i+1}^l \frac{1}{p_j})^2}\right) + \omega_l \leq \omega_l + \sum_{i=1}^{l-1} \omega_i \min\left(1, \frac{i}{(l-i)^2}\right) \\
& \leq \omega_{\max} \left(1 + \sum_{i=1}^{l-1} \min\left(1, \frac{i}{(l-i)^2}\right)\right) \stackrel{(b)}{\leq} 2\omega_{\max} \sqrt{l + \frac{1}{4}},
\end{aligned}$$

where (a) follows from Chebyshev's inequality; and (b) follows from Lemma 6. ■